



Faculté des Sciences Exactes et d'Informatique

Département de Mathématiques et informatique

Filière : Informatique

MÉMOIRE DE FIN D'ÉTUDES

pour l'obtention du diplôme de Master en Informatique

Option : Intelligence Artificielle et l'Internet des Objets

THÈME :

**La reconnaissance des émotions faciales en utilisant les
réseaux de neurones convolutifs**

Présenté par :

BOUTALBI Mohammed Iliass

Encadrante: Mme KAID SLIMANE Bouchra

Année Universitaire 2024-2025

Résumé :

La reconnaissance des émotions faciales (Face Expression Recognition – FER) constitue aujourd'hui un axe de recherche stratégique, tant par ses implications technologiques que par ses applications sociétales. Elle joue un rôle fondamental dans des domaines variés tels que la sécurité des accès, l'analyse de la satisfaction client, l'assistance personnalisée et le diagnostic en santé mentale. Le présent mémoire s'intéresse au développement d'un système de reconnaissance automatique des émotions à partir d'images faciales statiques, en s'appuyant sur les avancées récentes des réseaux de neurones à convolution profonde (Deep Convolutional Neural Networks – DCNN).

L'objectif principal de ce travail est de concevoir, entraîner et évaluer un modèle CNN capable d'identifier de manière fiable les émotions de base – telles que la joie, la tristesse, la colère ou la surprise – à partir d'images en niveaux de gris de faible résolution (48×48 pixels). L'approche adoptée intègre des techniques de prétraitement d'images, d'optimisation d'architecture de réseau et d'ajustement des paramètres d'apprentissage, afin d'améliorer la capacité du modèle à extraire des caractéristiques discriminantes pour la classification émotionnelle.

Les expérimentations menées sur des jeux de données spécialisés démontrent l'efficacité du modèle proposé pour la reconnaissance émotionnelle faciale. Les résultats obtenus soulignent non seulement la robustesse des réseaux convolutifs pour ce type de tâche, mais ouvrent également des perspectives prometteuses en matière d'interaction homme-machine, de systèmes embarqués et d'applications en temps réel. Ce mémoire présente en détail la méthodologie employée, l'analyse critique des résultats et propose des pistes d'amélioration pour des travaux futurs.

Mots-clés: Reconnaissance des émotions faciales (FER), Réseaux de neurones convolutifs (CNN), Réseaux de neurones à convolution profonde (DCNN), Apprentissage profond, Classification d'images.

Abstract

Facial Expression Recognition (FER) has emerged as a strategic area of research due to its profound technological significance and wide range of societal applications. It plays a critical role in fields such as access security, customer satisfaction analysis, personalized assistance, and mental health diagnostics. This thesis focuses on the development of an automatic emotion recognition system based on static facial images, leveraging the advancements of Deep Convolutional Neural Networks (DCNN).

The primary objective of this work is to design, train, and evaluate a CNN model capable of reliably identifying basic emotions—such as joy, sadness, anger, and surprise—from low-resolution grayscale images (48×48 pixels). The proposed approach integrates image preprocessing techniques, network architecture optimization, and fine-tuning of learning parameters to enhance the model's ability to extract discriminative features for emotion classification.

Experiments conducted on specialized datasets demonstrate the effectiveness of the developed model for facial emotion recognition. The results highlight not only the robustness of convolutional architectures for this task but also point to promising prospects for applications in human–machine interaction, embedded systems, and real-time environments. This thesis provides a detailed presentation of the adopted methodology, a critical analysis of the results obtained, and suggests potential improvements for future research.

Keywords: Facial Expression Recognition (FER), Convolutional Neural Networks (CNN), Deep Convolutional Neural Networks (DCNN), Deep Learning, Image Classification.

ملخص :

(Face Expression Recognition – FER) التعرف على الانفعالات الوجهية

يُعدّ اليوم محوراً استراتيجياً في مجال البحث، وذلك لما له من تأثيرات تكنولوجية وتطبيقات مجتمعية واسعة. فهو يلعب دوراً أساسياً في مجالات متعددة مثل أمن الوصول، تحليل رضا العملاء، المساعدة المخصصة، والتشخيص في مجال الصحة النفسية.

يركز هذا المذكرة على تطوير نظام للتعرف التلقائي على الانفعالات من الصور الوجهية الثابتة، (Deep Convolutional Neural Networks – DCNN) بالاعتماد على التقدم الحاصل في الشبكات العصبية الالتفافية العميقة. CNN الهدف الرئيسي من هذا العمل هو تصميم وتدريب وتقييم نموذج (DCNN – Deep Convolutional Neural Networks) على التعرف بدقة على الانفعالات الأساسية – مثل الفرح، الحزن، الغضب أو الدهشة – انطلاقاً من (صور رمادية منخفضة الدقة 48×48 بكسل).

تعتمد المنهجية المتبعة على تقنيات معالجة الصور المسبقة، وتحسين هيكلية الشبكة، وضبط معلمات التعلم، وذلك لتعزيز قدرة النموذج على استخراج السمات المميزة لتصنيف الانفعالات.

وقد أظهرت التجارب المجراة على مجموعات بيانات متخصصة فعالية النموذج المقترح في التعرف على الانفعالات الوجهية، حيث تؤكد النتائج المستخلصة ليس فقط على قوة الشبكات الالتفافية في مثل هذه المهام، بل تفتح أيضاً آفاقاً واعدة في مجالات التفاعل بين الإنسان والحاسوب، الأنظمة المدمجة، والتطبيقات الزمن الحقيقي.

يعرض هذا المذكرة بالتفصيل المنهجية المعتمدة، والتحليل النقدي للنتائج، كما يقترح سبل تحسين يمكن استغلالها في أعمال مستقبلية.

كلمات مفتاحية:

التعرف على الانفعالات الوجهية (FER)، الشبكات العصبية الالتفافية (CNN)، الشبكات العصبية الالتفافية العميقة (DCNN)، التعلم العميق، تصنيف الصور.

Dédicace

À mes chers parents

À ma famille

À mes amis

-Mohammed Iliass-

Remerciement

Je suis très reconnaissant pour cette opportunité de travailler sur ce projet et je tiens à exprimer mes sincères remerciements à tous ceux qui ont contribué à sa réussite.

Tout d'abord, je remercie **Allah** le tout-puissant de m'avoir donné la capacité de mener à bien ce projet et de m'avoir guidé tout au long de ce parcours.

J'adresse également mes remerciements les plus sincères à mes professeurs et encadrants universitaires, en particulier à **mon encadrante** Mme **KAID SLIMANE Bouchra**, pour ses conseils avisés et sa disponibilité tout au long de ce projet.

Je remercie également les membres du jury pour leur temps et leur engagement envers l'évaluation de mon travail.

Mes remerciements vont également à tous les membres du département et de la faculté pour la qualité de leur enseignement, leur dévouement et leur soutien constant tout au long de mon parcours universitaire.

Enfin, je remercie tous ceux qui m'ont soutenu et encouragé de près ou de loin, sans lesquels ce projet n'aurait pas été possible. Leur soutien et leurs encouragements ont été d'une grande aide tout au long de ce projet.

Je tiens à leur exprimer toute ma gratitude pour leur soutien et leur confiance.

Merci infiniment.

Table des Figures :

	Figure	Page
01	Expression faciale de la joie.	15
02	Expression faciale de la tristesse.	16
03	Expression faciale de la colère.	16
04	Expression faciale de la peur.	16
05	Expression faciale du dégoût.	17
06	Expression faciale de la surprise	17
07	Expression faciale de mépris.	18
08	Schéma des 68 points de repère faciaux (landmarks) utilisés en vision par ordinateur.	20
09	Exemple d'architecture CNN (couches conv + pooling + fully connected).	24
10	Illustration conceptuelle de la différence entre Machine Learning et Deep Learning	27
11	Schéma du fonctionnement d'un réseau de neurones multicouche avec rétropropagation	28
12	Évolution des performances sur le benchmark ImageNet grâce au Deep Learning	29
13	Architecture générale d'un réseau de neurones convolutif	31
14	Composants fonctionnels d'un CNN typique	32
15	Courbes des principales fonctions d'activation	33
16	Échantillons d'images de la base FER2013.	35
17	Diagramme des cas d'utilisation de l'application FER	37
18	Diagramme de classes de l'application FER	38
19	Diagramme de déploiement de l'application FER	39
20	Diagramme de séquence – Test de reconnaissance d'émotion en temps réel via webcam	39
21	Diagramme de séquence – Test de reconnaissance d'émotion à partir d'une image	40
22	Construction du modèle CNN profond	47
23	Importation des bibliothèques nécessaires à l'entraînement du modèle	52
24	Importation des bibliothèques nécessaires à l'évaluation du modèle	52
25	Importation des bibliothèques nécessaires à l'application FER	53
26	Entraînement du modèle	53
27	Chargement du modèle et génération des prédictions sur l'ensemble de test	54
28	Courbe d'apprentissage du modèle pour 30 époques	55
29	Courbe d'apprentissage du modèle pour 100 époques	56
30	Matrice de confusion du modèle	57
31	rapport de classification	57
32	Interface d'accueil de l'application	58
33	Interface d'accueil – Mode sombre	59

34	Interface de sélection des deux modes de détection	60
35	Interface du mode - Image statique	60
36	Chargement d'une image par l'utilisateur	61
37	Exemple de prédiction d'émotion sur une image statique	61
38	Interface du mode - Webcam temps réel	62
39	Lancement de la détection des émotions via webcam	62
40	Arrêt de la détection des émotions via webcam	63
41	Courbe d'apprentissage du modèle CNN	63
42	Matrice de confusion des émotions détectées	64
43	Rapport de classification : précision, rappel et F1-score	64

Liste des tableaux :

Tableau 1.1	Les sept émotions de base et leurs unités d'action principales (FACS)
Tableau 1.2	Domaines d'Applications de la FER
Tableau 2.1	Comparaison entre Machine Learning traditionnel et Deep Learning
Tableau 2.2	Récapitulatif des couches fondamentales d'un CNN.
Tableau 2.3	Comparaison des principales fonctions d'activation utilisées dans les réseaux de neurones.
Tableau 2.4	Comparaison des principales bases de données utilisées en FER
Tableau 2.5	Revue de travaux récents utilisant les CNN pour la FER
Tableau 3.1	Hyperparamètres choisis pour l'entraînement du modèle CNN

Liste des abréviations :

FER	Reconnaissance des émotions faciales
CNN	Réseaux de neurones convolutifs
DCNN	Réseaux de neurones à convolution profonde
ANN	Réseaux de neurones artificiels

Table de matière :

Résumé	2
Dédicace	5
Remerciement	6
Table des figures	7
Liste des tableaux	8
Liste des abréviations	8
Introduction Générale	11
Chapitre 01 : État de l'art lié à La Reconnaissance des émotions	13
1. Introduction	14
2. Les émotions humaines et les expressions faciales	14
2.1 Définition et importance des émotions humaines	14
2.2. Théories principales des émotions	15
2.3. Introduction aux 7 expressions faciales universelles	15
2.4. Les expressions faciales en communication	19
2.5. Codification des expressions : FACS et alternatives	19
3. Méthodes de reconnaissance faciale	20
3.1. Approches classiques	20
3.1.1. Méthodes géométriques	20
3.1.2. Approches basées sur l'apparence	20
3.2. Limites des approches traditionnelles	21
3.3. Nécessité du Deep Learning	22
4. Vision par ordinateur et intelligence artificielle appliquée	22
4.1. Vision par ordinateur	22
4.2. Intelligence artificielle appliquée à l'image	23
4.3. Introduction aux réseaux de neurones convolutifs (CNN)	23
5. Applications de la FER	24
6. Conclusion	24
Chapitre 02 : Réseaux de Neurones Convolutifs (CNN)	25
1. Introduction	26
2. Fondements de l'Apprentissage Profond	26
2.1. Machine Learning vers Deep Learning:	26
2.2. Fonctionnement général d'un réseau de neurones	28
2.3. Progrès récents et convergence entre théorie et pratique	29
3. Réseaux de Neurones Convolutifs (CNN)	30
3.1. Architecture d'un CNN	30
3.2. Couches principales d'un CNN	31
3.3. Fonctions d'activation	32
4. Application des CNN à la reconnaissance d'émotions	34
4.1. Principales bases de données utilisées (FER2013, CK+, etc.)	34
4.2. Travaux de recherche récents sur FER par CNN	35
5. Présentation d'UML	36

5.1 Définition	36
5.2 Avantages d'UML	37
5.3 Inconvénients d'UML	37
6.La démarche de notre modélisation	37
6.1. Modèle de cas d'utilisation	37
6.2. Modèle d'Analyse	37
6.3. Modèle de déploiement	38
6.4 Modèle de Test	39
7 .Conclusion	40
Chapitre 03 : Conception, implémentation et expérimentation	41
1. Introduction	42
2. Présentation générale de l'approche proposée	42
2.1 Objectif et stratégie choisie	42
2.2 Architecture générale du système	43
3. Préparation des données	44
3.1 Description du dataset utilisé	44
3.2 Prétraitements appliqués aux images	44
4. Conception du modèle CNN	45
4.1. Choix de l'architecture	45
4.2. Hyperparamètres et stratégie d'optimisation	47
5. Développement de l'interface utilisateur	48
5.1. Description des fonctionnalités principales	48
5.1.1. Mode reconnaissance en temps réel (webcam)	49
5.1.2. Mode classification à partir d'images	49
5.2 Intégration des éléments d'évaluation dans l'interface	49
5.3 Technologies utilisées (Python, TensorFlow/Keras, OpenCV, etc.)	50
6. Entraînement et validation du modèle	53
6.1. Métriques d'évaluation utilisées	54
6.2. Résultats obtenus	55
7. Déploiement du système	58
7.1. Détection des visages en temps réel	59
7.2. Prédiction de l'émotion (temps réel + image statique)	59
8. Analyse critique des résultats	64
8.1. Discussion sur les performances	65
9. Conclusion	67
Conclusion générale.....	68
Bibliographie.....	69

Introduction générale

La reconnaissance des émotions faciales (Face Expression Recognition – FER) s'impose aujourd'hui comme un domaine de recherche stratégique à l'interface de l'intelligence artificielle, de la vision par ordinateur et des sciences humaines. À l'ère de la digitalisation massive et de l'automatisation des interactions, la capacité des systèmes informatiques à interpréter les expressions émotionnelles humaines devient un levier essentiel pour améliorer l'accessibilité, renforcer la sécurité et personnaliser les expériences utilisateurs. Que ce soit dans les dispositifs de contrôle d'accès, l'analyse de la satisfaction client, l'assistance personnalisée ou le soutien en santé mentale, la reconnaissance des émotions ouvre des perspectives technologiques et sociétales majeures [1].

Dans ce contexte, les avancées récentes en apprentissage profond, notamment l'utilisation des réseaux de neurones à convolution (CNN), ont permis d'améliorer significativement la précision et la robustesse des systèmes de reconnaissance faciale. Grâce aux puissantes capacités d'extraction automatique de caractéristiques offertes par les CNN, il est désormais possible de concevoir des modèles capables de détecter et classifier des émotions complexes en temps réel ou à partir d'images statiques, même de faible résolution.

La motivation principale de cette recherche repose sur l'importance croissante de l'interprétation émotionnelle dans les interactions homme-machine, dans un monde où les dispositifs intelligents occupent une place centrale. La reconnaissance des émotions n'est plus seulement un défi technique, mais aussi un enjeu éthique, psychologique et social, nécessitant des solutions robustes, fiables et respectueuses des particularités humaines.

Ce travail vise ainsi à développer un système de reconnaissance automatique des émotions faciales basé sur l'apprentissage profond, avec un fonctionnement à **deux modes complémentaires** :

- **Mode 1 : reconnaissance des émotions faciales en temps réel** à partir d'une capture vidéo via une webcam.
- **Mode 2 : classification d'images statiques** chargées manuellement par l'utilisateur.

L'objectif de ce mémoire est de concevoir, entraîner et déployer un modèle CNN capable d'identifier avec précision des émotions de base (joie, tristesse, colère, surprise, peur, dégoût, neutralité) à partir de ces deux types d'entrée, tout en offrant une **interface utilisateur simple et intuitive**.

Pour atteindre cet objectif, plusieurs défis spécifiques sont abordés et qui sont les suivants :

- Concevoir une architecture CNN adaptée aux contraintes de la reconnaissance émotionnelle sur images de faible résolution.
- Mettre en œuvre des techniques de prétraitement et d'augmentation des données pour renforcer la robustesse du modèle.
- Optimiser les hyperparamètres d'apprentissage afin d'améliorer la précision sans provoquer de sur-apprentissage.
- Développer une interface conviviale permettant à la fois la reconnaissance en temps réel et la classification d'images statiques.

La problématique centrale à laquelle ce mémoire tente de répondre est le fait de concevoir et déployer un modèle CNN capable de reconnaître efficacement les émotions faciales en temps réel ou à partir d'images fournies par l'utilisateur, tout en assurant une interaction fluide via une interface adaptée.

Pour ce faire nous avons organisé ce mémoire en trois chapitres :

- **Chapitre 1 : État de l'art lié à La Reconnaissance des émotions**
Il présente les concepts fondamentaux liés aux émotions humaines, aux expressions faciales et aux approches traditionnelles de reconnaissance des émotions. Il expose également les principales méthodes utilisées dans ce domaine, ainsi que les limites des approches classiques.
- **Chapitre 2 : Réseaux de Neurones Convolutifs (CNN)**
Ce chapitre introduit en détail l'apprentissage profond et les réseaux de neurones convolutifs. Il présente également les fondements théoriques nécessaires à la compréhension du modèle développé dans ce travail.
- **Chapitre 3 : Conception, implémentation et expérimentation**
Ce dernier chapitre est consacré à la description de la méthodologie adoptée pour la conception du modèle CNN, son entraînement, ainsi que l'intégration dans une interface permettant la reconnaissance en temps réel et la classification à partir d'images. Les résultats expérimentaux obtenus sont analysés et discutés, en mettant en évidence les performances du système et les axes potentiels d'amélioration.

Chapitre 1 :État de l’art lié à La Reconnaissance des émotions

Chapitre 01 – État de l’art : La Reconnaissance des émotions

1. Introduction :

La reconnaissance des émotions faciales (FER) constitue un champ de recherche majeur de l’intelligence artificielle. Dans un monde de plus en plus automatisé, la capacité des systèmes informatiques à interpréter fidèlement les expressions émotionnelles humaines est cruciale pour plusieurs applications : sécurité des accès biométriques, évaluation de la satisfaction client, assistance et suivi en santé mentale, interfaces adaptatives en temps réel, etc. Cette diversité d’usages souligne l’importance de disposer d’outils fiables et performants pour la classification automatique des émotions faciales à partir d’images statiques ou de flux vidéo.

Ce premier chapitre a pour objectif de définir les concepts fondamentaux liés aux émotions et aux expressions faciales, de présenter les principales théories psychologiques, d’exposer les sept expressions universelles, et d’introduire les principales codifications (FACS, Action Units). Enfin, nous décrirons les limites des approches classiques et la nécessité de recourir aux techniques d’apprentissage profond.

2. Les émotions humaines et les expressions faciales :

2.1 Définition et importance des émotions humaines

Les émotions se définissent comme des états mentaux complexes, résultant de l’interaction entre :

1. Une **réponse physiologique** (rythme cardiaque, sudation, activation neuromusculaire).
2. Une **expérience subjective** (sentiment de joie, de peur ou de colère).
3. Une **manifestation comportementale** (posture, intonation vocale, expression faciale).

Ces trois dimensions s’articulent pour influencer notre prise de décision, la mémoire et l’interprétation sociale. Les travaux fondateurs d’Ekman et Friesen (1978) identifient sept émotions dites « de base » joie, tristesse, colère, peur,

dégoût, surprise et neutralité, qui se traduisent par des configurations faciales universelles, indépendamment des cultures [1]. Ces configurations, stables et répétables, constituent le socle de la reconnaissance automatique des émotions faciales (FER).

2.2. Théories principales des émotions

- **Ekman & Friesen (1978)** : postulat d'émotions « universelles » codifiées par des Action Units [1].
- **Plutchik (1980)** : modèle en « roue des émotions », huit émotions primaires en paires opposées illustrant les dynamiques de combinaison et d'intensité [2].
- **Russell (1980)** : approche circumplexe situant chaque émotion selon deux axes valence (positif vs négatif) et activation (bas vs haut) [3].

2.3. Introduction aux 7 expressions faciales universelles :

Paul Ekman et Wallace V. Friesen ont identifié les sept expressions faciales universelles, observées de manière constante à travers différentes cultures et sociétés. Leur universalité suggère une base biologique commune à tous les êtres humains.

Chaque émotion se manifeste par des configurations spécifiques des muscles faciaux, appelées unités d'action (Action Units, AU), décrites dans le Facial Action Coding System (FACS). Par exemple :

- **Joie** : AU6 (contraction des muscles autour des yeux) et AU12 (élévation des coins de la bouche).



Figure 1.1 - Expression faciale de la joie. [1]

- **Tristesse** : AU1 (élévation des sourcils internes) et AU15 (abaissement des coins de la bouche).

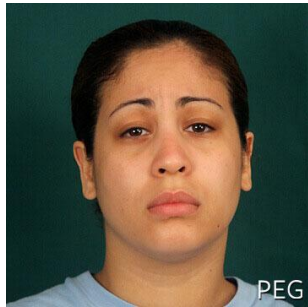


Figure 1.2 - Expression faciale de la tristesse. [1]

- **Colère** : AU4 (froncement des sourcils) et AU7 (resserrement des paupières).

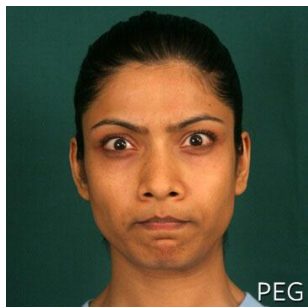


Figure 1.3 - Expression faciale de la colère. [1]

- **Peur** : AU1+2 (élévation des sourcils) et AU5 (ouverture des paupières supérieures).

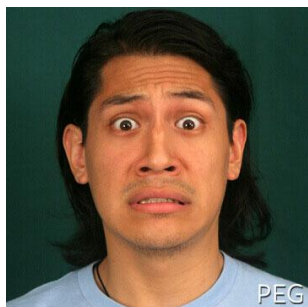


Figure 1.4 - Expression faciale de la peur. [1]

- **Dégoût** : AU9 (plissement du nez) et AU10 (élévation de la lèvre supérieure).

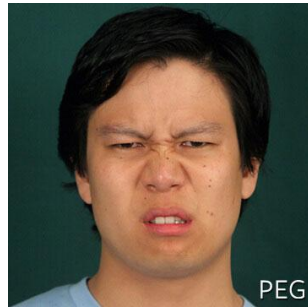


Figure 1.5 - Expression faciale du dégoût. [1]

- **Surprise** : AU1+2 (élévation des sourcils) et AU26 (abaissement de la mâchoire).

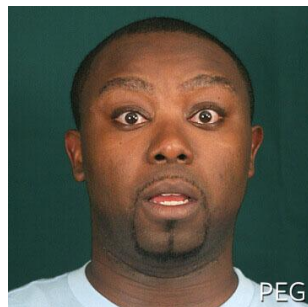


Figure 1.6 - Expression faciale de la surprise. [1]

- **Mépris** : élévation unilatérale du coin de la bouche.



Figure 1.7 - Expression faciale de mépris. [1]

Ces expressions jouent un rôle crucial dans la communication non verbale, permettant de transmettre des émotions et des intentions sans recours à la parole. Elles sont également fondamentales dans des domaines tels que la psychologie, la sécurité, et l'interaction homme-machine.

Tableau 1.1 – Les sept émotions de base et leurs unités d’action principales (FACS)

Émotion	AU principales	Description brève
Joie	AU6 (zglissement des joues), AU12 (zygomatic major)	Sourire appuyé, yeux plissés
Tristesse	AU1 (inner brow raiser), AU15 (lip corner depressor)	Lèvre inférieure tombante
Colère	AU4 (brow lowerer), AU7 (lid tightener)	Regards perçants, mâchoire tendue

Peur	AU1+2 (brow raiser), AU5 (upper lid raiser)	Yeux très ouverts, bouche entrouverte
Dégoût	AU9 (nose wrinkler), AU10 (upper lip raiser)	Nez plissé, lèvre retroussée
Surprise	AU1+2 (brow raiser), AU26 (jaw drop)	Yeux écarquillés, mâchoire descendue
Neutralité	Aucune AU marquante	Visage détendu sans hyperfonctionnement

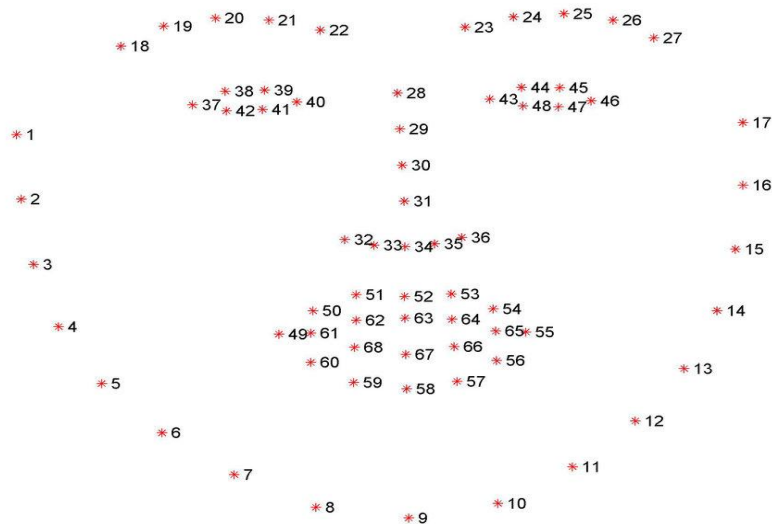
2.4. Les expressions faciales en communication

Les expressions faciales constituent un canal majeur de la communication non verbale :

1. **Signal adaptatif** : informer autrui d'un danger, d'une satisfaction ou d'une menace.
2. **Cognition sociale** : perception rapide (< 200 ms) d'une émotion impactant instantanément le traitement cognitif et le comportement social [4].
3. **Interopérabilité culturelle** : validité interculturelle des émotions de base, justifiant la pertinence de la FER universelle.

2.5. Codification des expressions : FACS et alternatives

- **Facial Action Coding System (FACS)** : codification exhaustive des micro- expressions via des AU, élaborée par Ekman & Friesen [1].
- **Landmarks faciaux** : extraction de 68 points (Dlib) ou plus (MediaPipe) pour calculer les distances et les angles.
- **Modèles paramétriques (AAM, ASM)** : combinaison de forme et d'apparence dans un espace réduit.



- **Figure 1.8:** Schéma des 68 points de repère faciaux (landmarks) utilisés en vision par ordinateur. [5]

3. Méthodes de reconnaissance faciale :

3.1. Approches classiques:

3.1.1. Méthodes géométriques:

Les méthodes géométriques reposent sur l'analyse des relations spatiales entre des points clés du visage, tels que les yeux, le nez, la bouche et les sourcils. Ces approches extraient des caractéristiques telles que les distances inter-oculaires, les angles formés par les traits du visage ou la hauteur du sourire. Bien que ces méthodes soient intuitives et peu coûteuses en calcul, elles présentent une sensibilité accrue aux variations d'orientation, d'éclairage et d'expression faciale, limitant ainsi leur robustesse dans des conditions réelles.

3.1.2. Approches basées sur l'apparence :

Les méthodes basées sur l'apparence exploitent la texture et les motifs visuels du visage pour en extraire des descripteurs discriminants. Parmi les techniques les plus répandues figurent :

- **Local Binary Patterns (LBP)** : Cette méthode encode la texture locale en comparant chaque pixel à ses voisins, générant ainsi un histogramme caractéristique de la région analysée. Elle est particulièrement efficace pour capturer les micro-textures du visage, telles que les rides ou les pores [6].
- **Histogram of Oriented Gradients (HOG)** : HOG divise l'image en cellules et calcule l'orientation des gradients de chaque pixel, produisant ainsi un histogramme des orientations dominantes. Cette technique est robuste aux variations d'éclairage et est largement utilisée pour la détection d'objets, y compris les visages [7].
- **Filtres de Gabor** : Inspirés du système visuel humain, les filtres de Gabor capturent les informations de fréquence et d'orientation, permettant une représentation efficace des textures faciales. Ils sont particulièrement utiles pour détecter des motifs spécifiques tels que les rides ou les contours des traits du visage [8].

Ces méthodes ont démontré leur efficacité dans des environnements contrôlés, mais leur performance peut se dégrader en présence de variations significatives des conditions d'acquisition.

3.2. Limites des approches traditionnelles

Malgré leurs contributions significatives, les approches classiques présentent plusieurs limitations :

- **Sensibilité aux conditions d'acquisition** : Les variations d'éclairage, d'orientation de la tête et d'expression faciale peuvent affecter la fiabilité des descripteurs extraits.
- **Extraction manuelle des caractéristiques** : La nécessité de définir manuellement les descripteurs pertinents peut entraîner une perte d'information et limiter la généralisation des modèles.
- **Incapacité à capturer des représentations hiérarchiques** : Les méthodes traditionnelles peinent à modéliser les relations complexes entre les différentes caractéristiques du visage, réduisant ainsi leur capacité à distinguer des individus dans des conditions variées.

Ces limitations ont motivé la recherche de solutions plus robustes et adaptatives, conduisant à l'émergence des approches basées sur l'apprentissage profond.

3.3. Nécessité du Deep Learning:

L'avènement des réseaux de neurones profonds, en particulier les Convolutional Neural Networks (CNN), a révolutionné le domaine de la reconnaissance faciale. Contrairement aux méthodes traditionnelles, les CNN apprennent automatiquement des représentations hiérarchiques des données, capturant à la fois les caractéristiques locales et globales du visage.

Parmi les avantages des CNN, on peut citer :

- **Robustesse aux variations** : Les CNN sont capables de généraliser efficacement face aux variations d'éclairage, d'orientation et d'expression faciale, améliorant ainsi la précision de la reconnaissance [9].
- **Extraction automatique des caractéristiques** : Les CNN éliminent le besoin d'une extraction manuelle des descripteurs, apprenant directement à partir des données brutes les caractéristiques les plus discriminantes.
- **Capacité à modéliser des relations complexes** : Grâce à leur architecture profonde, les CNN peuvent capturer des relations non linéaires entre les différentes caractéristiques du visage, améliorant ainsi la différenciation entre individus.

Des architectures telles que VGG-Face, FaceNet et DeepFace ont démontré des performances remarquables dans des benchmarks standardisés, surpassant les approches traditionnelles et établissant de nouveaux standards dans le domaine de la reconnaissance faciale [10].

4. Vision par ordinateur et intelligence artificielle appliquée:

4.1. Vision par ordinateur :

La vision par ordinateur est un domaine de l'intelligence artificielle (IA) qui permet aux machines d'interpréter et de comprendre les informations visuelles issues du monde réel, telles que les images et les vidéos. Elle vise à reproduire les capacités de perception humaine en dotant les systèmes informatiques de la faculté de voir et d'analyser leur environnement visuel. Les tâches fondamentales de la vision par ordinateur incluent :

- **Détection d'objets** : Identification et localisation des objets présents dans une image ou une vidéo.

- **Segmentation d'images** : Division d'une image en segments significatifs pour isoler des objets ou des régions d'intérêt.
- **Reconnaissance de formes** : Identification de motifs ou de structures spécifiques au sein d'une image.
- **Suivi d'objets** : Suivi du mouvement d'objets à travers une séquence d'images ou une vidéo.

Ces techniques sont largement utilisées dans divers domaines tels que la sécurité, la médecine, l'automobile et la robotique, permettant des applications allant de la reconnaissance faciale à la conduite autonome.

4.2. Intelligence artificielle appliquée à l'image:

L'intelligence artificielle appliquée à l'image englobe l'utilisation d'algorithmes avancés pour analyser, interpréter et générer des contenus visuels. Elle s'appuie sur des techniques telles que les réseaux de neurones convolutifs (CNN), les réseaux antagonistes génératifs (GAN) et les modèles de diffusion pour accomplir diverses tâches :

- **Génération d'images** : Création d'images réalistes à partir de descriptions textuelles ou d'autres entrées, comme le font des outils tels que DALL-E, Midjourney ou Bing Image Creator [11].
- **Amélioration et restauration** : Amélioration de la qualité des images, suppression du bruit, restauration de photos anciennes ou endommagées.
- **Transfert de style artistique** : Application du style d'une image (par exemple, une peinture célèbre) à une autre image, permettant des créations artistiques innovantes [12].
- **Analyse médicale** : Détection et diagnostic assistés par IA dans des images médicales, comme la radiographie ou l'IRM, pour identifier des anomalies ou des pathologies [13].
- **Surveillance et sécurité** : Analyse en temps réel des flux vidéo pour la détection d'activités suspectes ou la reconnaissance faciale dans des contextes de sécurité [14].

4.3. Introduction aux réseaux de neurones convolutifs (CNN):

Le CNN se compose de couches convolutionnelles, pooling, activation ReLU, couches denses (Figure 1.9):

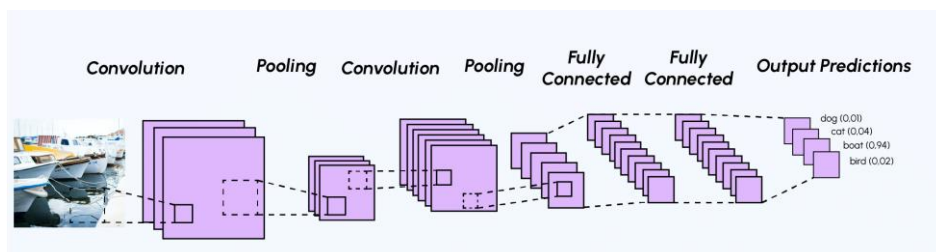


Figure 1.9 – Exemple d’architecture CNN (couches conv + pooling + fully connected).

5. Applications de la FER:

Tableau 1.2 : Domaines d’Applications de la FER

Domaine	Usage principal
Sécurité	Contrôle d’accès biométrique, détection de menace
Santé mentale	Diagnostic précoce de troubles, suivi du bien- être émotionnel
Marketing	Analyse de satisfaction, UX testing
Interfaces	Agents virtuels empathiques, robots sociaux

Des études récentes démontrent que l’intégration de la FER dans les systèmes temps réel (webcam, vidéo- stream) permet d’adapter dynamiquement l’environnement à l’utilisateur [15, 16].

6. Conclusion :

Dans ce chapitre nous avons présenté le cadre théorique de la reconnaissance des émotions faciales en citant les concepts d’émotion et codification faciale, nous avons donné les limites des méthodes traditionnelles, ce qui a conduit à une ouverture vers le deep learning, le Chapitre qui suit détaillera la méthodologie propre au développement d’un modèle CNN pour la classification d’émotions en flux webcam et à partir d’images statiques.

Chapitre 2- Réseaux de Neurones Convolutifs (CNN)

1. Introduction :

L'intelligence artificielle, et plus particulièrement l'apprentissage profond (Deep Learning), a révolutionné de nombreux domaines, allant de la vision par ordinateur à la reconnaissance vocale. Dans ce chapitre, nous nous intéressons aux fondements théoriques et techniques qui sous-tendent notre projet de reconnaissance des expressions faciales à l'aide des réseaux de neurones convolutifs (CNN). Nous commençons par une présentation générale de l'apprentissage profond et de ses principes de fonctionnement, puis nous explorons les CNN, qui sont aujourd'hui les architectures les plus performantes pour le traitement des images. Ensuite, nous abordons leur application spécifique à la reconnaissance des émotions à partir d'images faciales.

Dans un second temps, nous introduisons le langage de modélisation UML qui permet de représenter de manière claire et structurée l'architecture fonctionnelle et logicielle de notre application. Nous terminons par la démarche suivie pour la modélisation de notre système, en mettant en avant les principaux diagrammes utilisés dans ce contexte.

2. Fondements de l'Apprentissage Profond :

L'apprentissage profond (Deep Learning) constitue une sous-branche du Machine Learning qui a récemment connu une croissance exponentielle, tant en termes de recherche que d'applications pratiques. Il repose sur l'utilisation de réseaux de neurones artificiels à plusieurs couches (dits « profonds »), capables d'extraire automatiquement des représentations hiérarchiques et abstraites à partir de données brutes. Avant d'explorer en détail les CNN, il est nécessaire d'établir les fondements conceptuels de l'apprentissage profond, notamment en distinguant les paradigmes classiques du Machine Learning, en exposant le fonctionnement général d'un réseau de neurones artificiel, et en précisant les mécanismes qui permettent l'optimisation de ces modèles.

2.1. Machine Learning vers Deep Learning:

Le Machine Learning (ML), ou apprentissage automatique, désigne un ensemble de techniques algorithmiques permettant à une machine d'apprendre des modèles à partir de données, sans être explicitement programmée pour accomplir une tâche spécifique [18]. Il se décline en

plusieurs paradigmes : apprentissage supervisé, non supervisé, semi-supervisé et par renforcement.

Le Deep Learning s’inscrit dans ce cadre, mais s’en distingue par la profondeur de ses modèles (Figure 2.1). Alors que les approches classiques du ML s’appuient souvent sur une ingénierie manuelle des caractéristiques (feature engineering), les architectures profondes apprennent directement à partir des données brutes, réduisant ainsi la dépendance aux connaissances a priori [19].

Tableau 2.1 – Comparaison entre Machine Learning traditionnel et Deep Learning

Critère	Machine Learning traditionnel	Deep Learning
Nécessité d'extraction manuelle des features	Oui	Non (extraction automatique)
Volume de données requis	Modéré	Très élevé
Capacité à modéliser la complexité	Limité	Élevée grâce aux multiples couches
Interprétabilité	Relativement bonne	Moins interprétable
Temps d'entraînement	Court à modéré	Long (nécessite du calcul intensif)

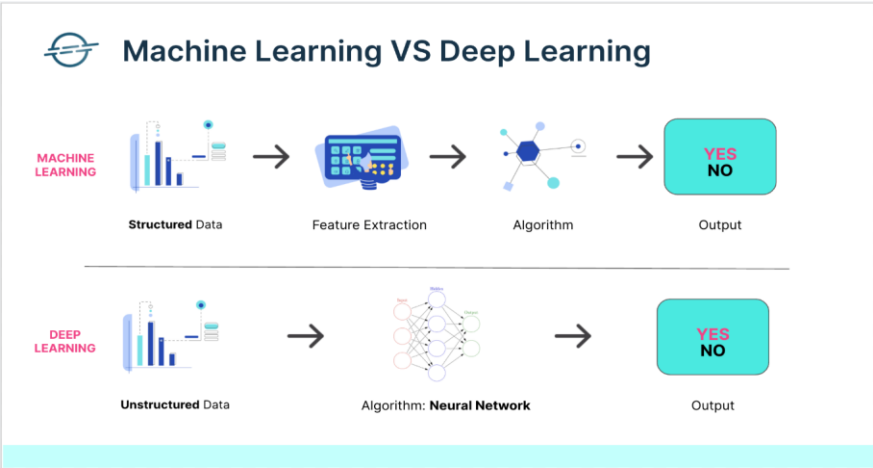


Figure 2.1– Illustration conceptuelle de la différence entre Machine Learning et Deep Learning

2.2. Fonctionnement général d'un réseau de neurones

Un réseau de neurones artificiels (ANN) est composé d'unités de calcul appelées neurones, organisées en couches successives : une couche d'entrée, une ou plusieurs couches cachées, et une couche de sortie. Chaque neurone effectue une opération affine (pondération et biais), suivie d'une fonction d'activation non linéaire, comme l'indique la Formule (2.1) :

$$y = f(w \cdot x + b) \quad (2.1)$$

Où x est le vecteur d'entrée, w le vecteur de poids, b le biais, et f une fonction d'activation.

Le processus d'apprentissage consiste à ajuster les poids (w) et les biais (b) pour minimiser une fonction de perte (*loss function*), à l'aide d'une méthode d'optimisation telle que la descente de gradient stochastique (SGD) [20]. L'algorithme de rétropropagation (backpropagation) permet de calculer efficacement les gradients à chaque couche via la chaîne de dérivations (Figure 2.2).

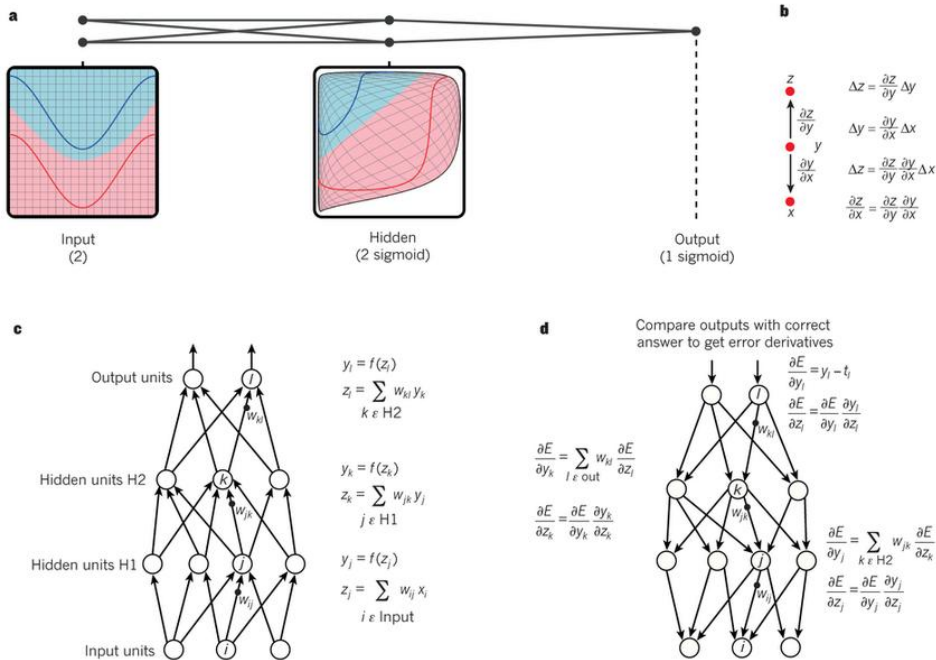


Figure 2.2 – Schéma du fonctionnement d'un réseau de neurones multicouche avec rétropropagation. [23]

L'empilement de plusieurs couches non linéaires permet au réseau de construire progressivement des représentations de plus en plus abstraites : de simples bords dans les premières couches, à des formes complexes, voire des objets entiers, dans les couches profondes [21].

2.3. Progrès récents et convergence entre théorie et pratique

Les avancées matérielles (GPU, TPU), l'augmentation exponentielle des volumes de données et le perfectionnement des architectures ont permis aux réseaux de neurones profonds de dépasser les méthodes classiques dans de nombreux domaines : vision par ordinateur, traitement automatique du langage naturel, bioinformatique, etc. Les performances obtenues sur des benchmarks tels qu'ImageNet ont catalysé une adoption massive des réseaux convolutionnels, en particulier dans la tâche de reconnaissance faciale, comme le montre la Figure 2.3, adaptée d'une synthèse de résultats publiés entre 2010 et 2017.

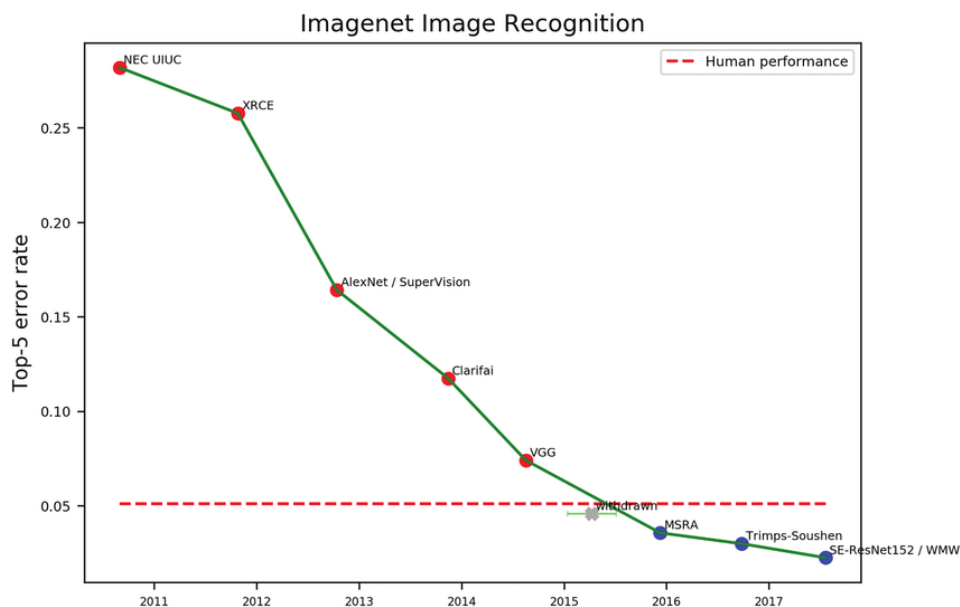


Figure 2.3 – Évolution des performances sur le benchmark ImageNet grâce au Deep Learning (2010–2017). [32]

Ainsi, l'apprentissage profond représente une rupture épistémologique majeure en intelligence artificielle, en ce qu'il repose sur une autonomie d'apprentissage fondée sur la hiérarchisation des représentations. Ce paradigme, que nous explorons plus en détail dans les sections suivantes, constitue le socle des systèmes modernes de reconnaissance d'émotions par vision artificielle.

3. Réseaux de Neurones Convolutifs (CNN):

Les Réseaux de Neurones Convolutifs (CNN) constituent l'un des piliers de l'apprentissage profond, particulièrement adaptés au traitement des données visuelles telles que les images. Leur succès est dû à leur capacité à extraire automatiquement des caractéristiques discriminantes tout en conservant la structure spatiale des données d'entrée.

3.1. Architecture d'un CNN: [23]

Un CNN typique repose sur une succession de couches spécialisées organisées en pipeline, avec en entrée une image (souvent représentée par un tenseur tridimensionnel : hauteur, largeur, canaux), et en sortie une prédiction (classe, régression, etc.).

L'architecture se compose généralement de trois blocs fonctionnels :

- **Le bloc d'extraction de caractéristiques** : constitué de couches convolutionnelles, éventuellement suivies de fonctions d'activation et de couches de pooling (sous-échantillonnage).
- **Le bloc de mise à plat (flatten)** : transforme les cartes de caractéristiques 2D en un vecteur 1D.
- **Le classifieur** : enchaînement de couches entièrement connectées (fully connected) produisant la sortie finale via une fonction d'activation (ex. softmax).

La Figure 2.4 illustre de manière schématique une architecture standard de CNN, mettant en évidence l'enchaînement logique des couches.

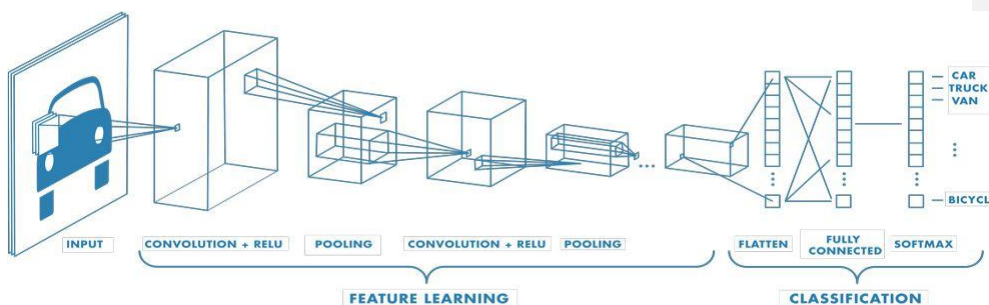


Figure 2.4 – Architecture générale d'un réseau de neurones convolutif.

Cette hiérarchie permet au réseau de capturer progressivement des motifs simples (bords, textures) dans les premières couches, puis des formes complexes et des entités sémantiques dans les couches profondes.

3.2. Couches principales d'un CNN : [24]

Un CNN repose sur des types de couches spécifiques qui assurent différentes fonctions au sein de l'architecture :

Tableau 2.2 – Récapitulatif des couches fondamentales d'un CNN.

Type de couche	Rôle	Détails techniques
Convolution	Extraire des motifs visuels locaux	Noyaux (kernels) glissés sur l'image
Activation (ReLU, etc.)	Introduire la non-linéarité	$f(x)=\max(0,x)$
Pooling (max, average)	Réduire la dimensionnalité tout en conservant l'information locale	MaxPooling 2x2, etc.
Flatten	Vectoriser les cartes de caractéristiques pour les couches denses	Passage de 3D à 1D
Fully Connected (Dense)	Apprentissage global / décision finale	Produit matriciel et biais, suivi d'une activation
Dropout (optionnel)	Régularisation pour éviter l'overfitting	Désactive aléatoirement des neurones pendant l'entraînement

Le Figure 2.5 illustre ces différentes couches dans une architecture typique.

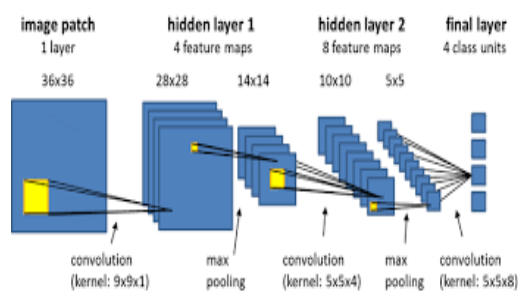


Figure 2.5 – Composants fonctionnels d'un CNN typique.

Chaque couche joue un rôle fondamental dans la transformation progressive de l'image en représentation abstraite permettant la prise de décision en fin de réseau.

3.3. Fonctions d'activation :

Les fonctions d'activation sont essentielles dans les réseaux de neurones, car elles permettent de capturer des relations non-linéaires. Sans elles, un réseau profond ne serait qu'une combinaison linéaire de ses entrées, donc limité en expressivité.

Parmi les fonctions les plus utilisées dans les CNN, on distingue :

Tableau 2.3- Comparaison des principales fonctions d'activation utilisées dans les réseaux de neurones.

Fonction	Formule mathématique	Avantages	Inconvénients
ReLU	$f(x)=\max\{0,x\}$	Simple, efficace, rapide à converger	Neurones morts (Dead ReLU)
Sigmoïde	$f(x)=1/(1+e^{-x})$	Interprétation probabiliste	Saturation, vanishing gradients
tanh	$f(x)=(e^x-e^{-x})/(e^x+e^{-x})$	Centre les données	Vanishing gradients
Leaky ReLU	$f(x)=x$ si $x>0$, sinon αx	Évite les neurones morts	Dépend du choix de α
Softmax	$\sigma(z_i) = e^{z_i} / \sum_j e^{z_j}$	Produits de probabilités pour multi-classes	Réservée à la couche de sortie

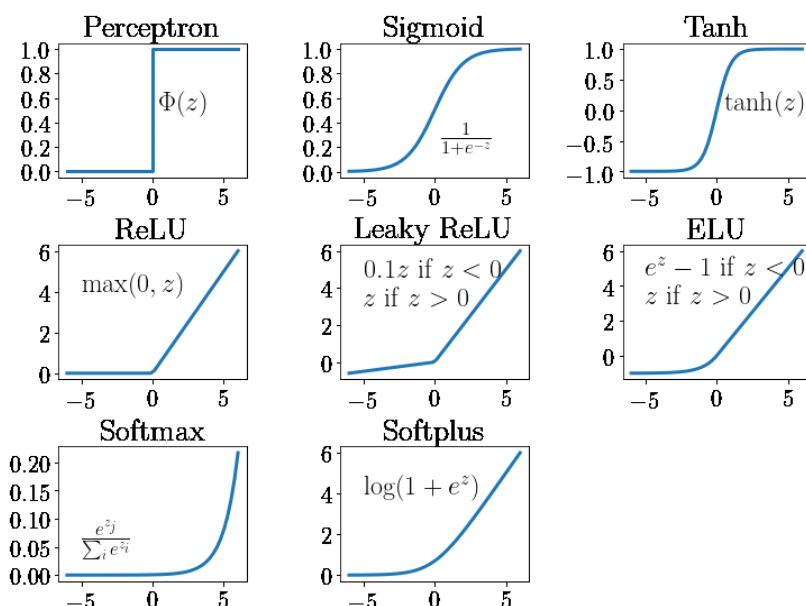


Figure 2.6 – Courbes des principales fonctions d’activation.

Ces fonctions conditionnent la dynamique d’apprentissage. ReLU est la plus répandue dans les couches cachées, tandis que Softmax est privilégiée en sortie pour les tâches de classification multiclasse.

4. Application des CNN à la reconnaissance d’émotions:

La reconnaissance d’émotions faciales (FER) est un domaine de recherche intensément exploré, combinant la vision par ordinateur et l’apprentissage profond. Les réseaux de neurones convolutifs (CNN) ont permis des avancées majeures dans cette tâche en automatisant l’extraction de caractéristiques discriminantes à partir d’images faciales. Cette section présente les principales bases de données utilisées dans ce domaine, ainsi qu’une revue des travaux récents fondés sur les CNN.

4.1. Principales bases de données utilisées (FER2013, CK+, etc.):

Les performances des algorithmes de reconnaissance d’émotions dépendent fortement de la qualité et de la diversité des données utilisées pour l’entraînement. Plusieurs bases de données de référence sont utilisées dans la littérature, chacune

présentant des spécificités:

Tableau 2.4 – Comparaison des principales bases de données utilisées en FER

Nom de la base	Nombre d'images	Émotions annotées	Format	Année	Remarques
FER2013	35 887	7 (colère, dégoût, peur, joie, tristesse, surprise, neutre)	48×48 (N&B)	2013	Très utilisée pour entraîner des CNN légers [25]
CK+ (Extended Cohn-Kanade)	593 vidéos (123 sujets)	7 émotions principales	Vidéos/ images	2010	Labels codés en FACS, expressions posées [26]
JAFFE	213 images	7 émotions faciales	256×256 (N&B)	1998	Expressions faciales de femmes japonaises
AffectNet	1 000 000+	Émotions continues et discrètes	Couleur	2017	Grande diversité, émotions valence-arousal [27]



Figure 2.7 – Échantillons d’images de la base FER2013.

La base **FER2013**, proposée dans le cadre du challenge ICML 2013, est probablement la plus populaire pour l’entraînement rapide de CNN grâce à son format simple (images 48x48, noir et blanc). La base **CK+**, bien que de plus petite taille, est souvent utilisée pour des tâches de validation, notamment dans les études de reconnaissance temporelle ou FACS.

4.2. Travaux de recherche récents sur FER par CNN:

Plusieurs travaux récents ont exploré l’application des CNN à la tâche de FER. L’approche typique consiste à concevoir un réseau convolutif profond capable d’extraire des motifs pertinents, tout en étant robuste aux variations d’éclairage, d’orientation ou d’identité.

Tableau 2.5 – Revue de travaux récents utilisant les CNN pour la FER

Étude	Méthodologie	Base utilisée	Précision (%)
Mollahosseini et al. (2016) [28]	Deep CNN à 6 couches avec dropout + ReLU	AffectNet	66.4
Yu et Zhang (2015) [25]	CNN simple avec 3 conv + 2 dense	FER2013	71.2

Hasani et Mahoor (2017) [29]	3D-CNN avec classification spatio-temporelle	CK+	93.2
Lopes et al. (2017) [30]	CNN léger pré-entraîné + fine-tuning	FER2013	72.4

Ces résultats montrent que la reconnaissance des émotions bénéficie significativement de l'apprentissage profond. L'utilisation de couches convolutives profondes, accompagnée de techniques telles que le dropout, la normalisation ou le data augmentation, améliore notablement les performances. En outre, l'utilisation de réseaux pré-entraînés (par exemple VGG ou ResNet) s'est avérée efficace, surtout pour les petites bases comme JAFFE ou CK+.

Certaines architectures récentes intègrent également des mécanismes d'attention ou des modules temporels (LSTM, GRU) pour capter l'évolution des émotions dans des séquences vidéo [29].

5. Présentation d'UML :

5.1 Définition :

UML c'est l'acronyme anglais pour « Unified Modeling Language ». On le traduit par « Langage de modélisation unifié ». La notation UML est un langage visuel constitué d'un ensemble de schémas, appelés des diagrammes, qui donnent chacun une vision différente du projet à traiter. UML nous fournit donc des diagrammes pour représenter le logiciel à développer : son fonctionnement, sa mise en route, les actions susceptibles d'être effectuées par le logiciel, etc [31]

5.2 Avantages d'UML :

Il possède une sémantique propre et une syntaxe composée de graphe. soit dans l'inférence).

- Permet de construire plusieurs modèles.
- Langage universel.
- Facilité de traduire un texte plusieurs formes (diagrammes).

5.3 Inconvénients d'UML :

- Apprentissage nécessaire (période d'adaptation) .
- Ne constitue qu'une étape (n'est pas à l'origine des concepts objets).

6.La démarche de notre modélisation :

Nous allons présenter les différents diagrammes des modèles suivant :

- Un modèle de cas d'utilisation.
- Un modèle d'analyse.
- Un modèle de déploiement
- Un modèle de test.

6.1. Modèle de cas d'utilisation :

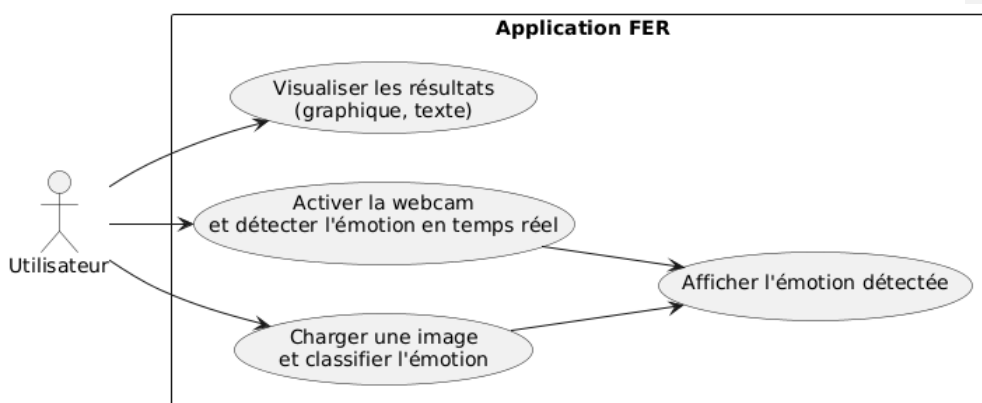


Figure 2.8- Diagramme des cas d'utilisation de l'application FER

6.2. Modèle d'Analyse :

6.2.1 Diagramme de classe:

Diagramme de classes - Modèle d'analyse de l'application FER

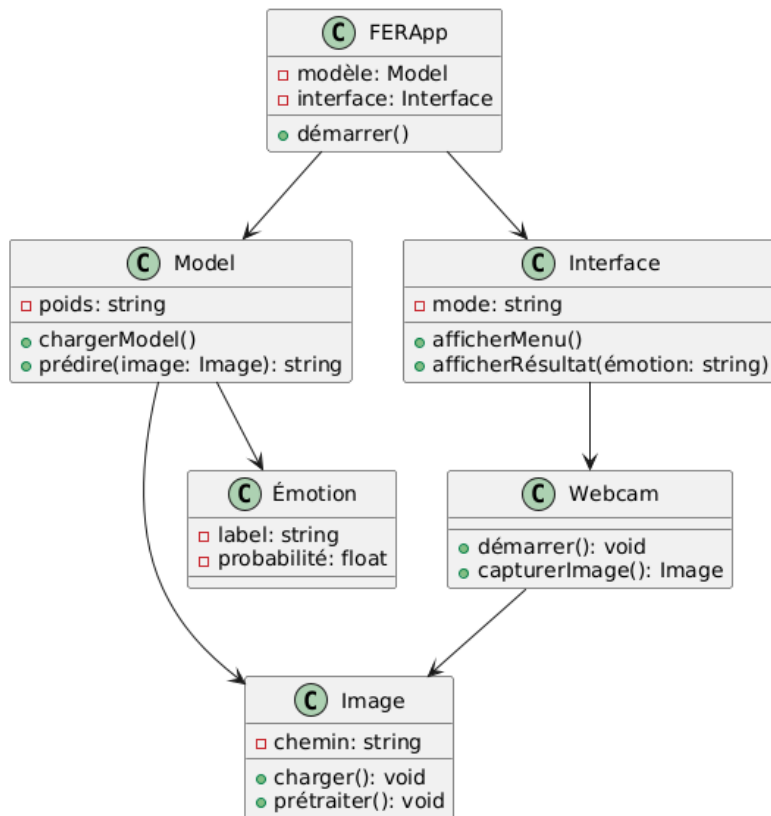


Figure 2.9- Diagramme de classes de l'application FER

6.3. Modèle de déploiement :

6.3.1 Diagramme de déploiement:

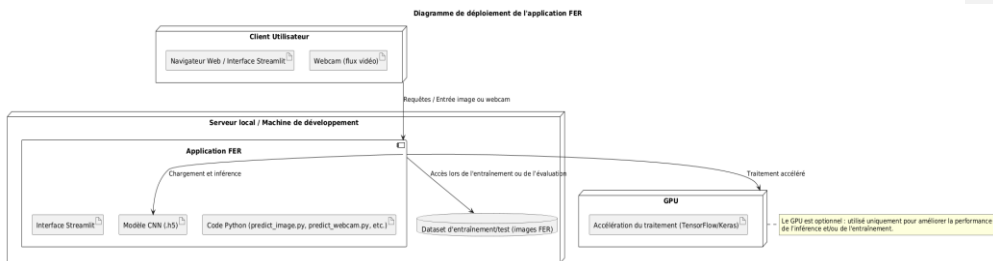


Figure 2.10- Diagramme de déploiement de l'application FER

6.4 Modèle de Test:

6.4.1 Diagrammes de Séquence :

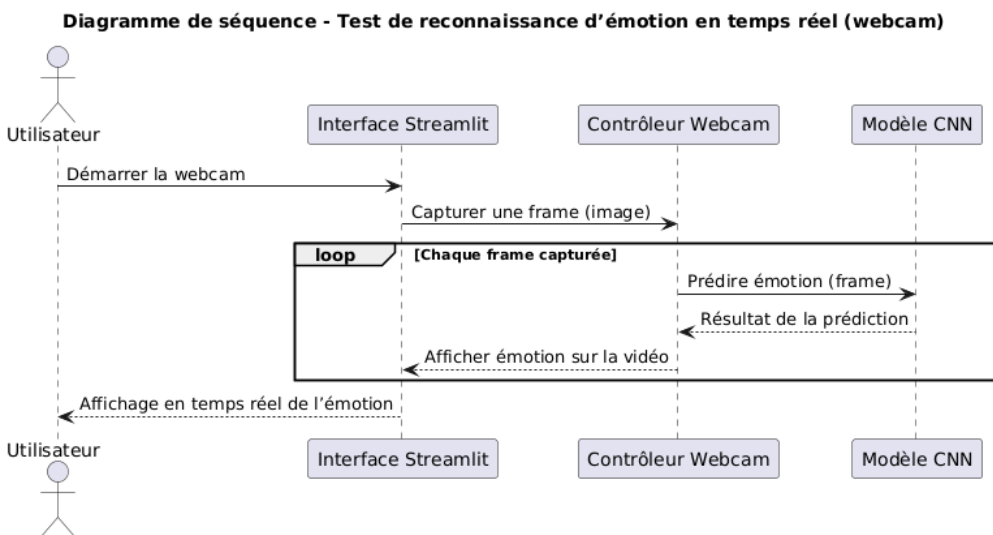


Figure 2.11 - Diagramme de séquence – Test de reconnaissance d'émotion en temps réel via webcam

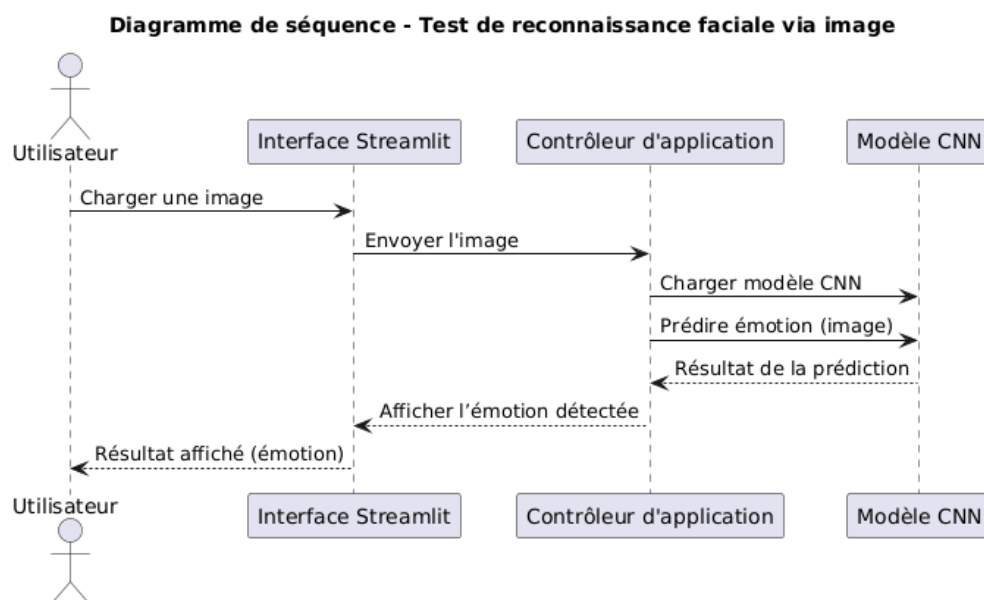


Figure 2.12 - Diagramme de séquence – Test de reconnaissance d’émotion à partir d’une image

7. Conclusion :

Ce chapitre nous a permis de poser les bases théoriques nécessaires à la compréhension de notre projet. Nous avons présenté les concepts clés de l’apprentissage profond, avec un focus particulier sur les réseaux de neurones convolutifs, leur architecture, leur fonctionnement et leur pertinence pour la reconnaissance d’émotions faciales. Nous avons également montré comment ces modèles peuvent être intégrés dans des applications concrètes de vision par ordinateur.

Enfin, nous avons introduit le langage UML en tant qu’outil de modélisation et décrit la démarche adoptée pour structurer notre application à travers différents diagrammes. Ces éléments préparent le terrain pour le chapitre suivant, dans lequel nous détaillerons la conception et la réalisation pratique de notre système de reconnaissance d’émotions faciales basé sur les CNN.

Chapitre 3-

Conception, implémentation et expérimentation

1. Introduction :

Mis en forme : Police :14 pt, Gras, Police de script complexe :14 pt, Gras

Après avoir établi les fondements théoriques de notre projet dans le chapitre précédent, nous abordons ici la phase de concrétisation de notre système de reconnaissance d'expressions faciales basé sur les réseaux de neurones convolutifs (CNN). Ce chapitre détaille l'ensemble du processus de mise en œuvre, depuis la formulation de l'approche adoptée jusqu'à l'analyse critique des résultats obtenus.

Nous commençons par présenter la stratégie générale de développement et l'architecture du système, avant de décrire les étapes clés du traitement des données, de la conception du modèle, et de la création de l'interface utilisateur. L'entraînement du modèle est ensuite présenté à travers les métriques d'évaluation utilisées et les résultats mesurés. Nous abordons également le processus de déploiement du système dans ses deux modes de fonctionnement : reconnaissance en temps réel via webcam et classification à partir d'images statiques.

Enfin, nous procédons à une discussion approfondie sur les performances observées, les limites identifiées, ainsi que les perspectives d'amélioration envisageables. Cette démarche rigoureuse vise à fournir une vision complète et cohérente du cycle de vie de notre système FER, de la conception initiale à son évaluation expérimentale.

2. Présentation générale de l'approche proposée

2.1 Objectif et stratégie choisie

Mis en forme : Police :14 pt, Police de script complexe :14 pt

L'objectif principal de notre projet est de concevoir un système capable de reconnaître les émotions faciales en temps réel ou à partir d'images statiques, en s'appuyant sur la puissance des réseaux de neurones convolutifs (CNN). Pour atteindre cet objectif, nous avons opté pour une approche modulaire et itérative reposant sur les étapes clés suivantes :

- La préparation et le prétraitement des données issues de bases d'images annotées d'expressions faciales.
- La conception et l'entraînement d'un modèle CNN adapté à la tâche de classification des émotions.
- Le développement d'une interface utilisateur intuitive, facilitant l'utilisation du système pour des images locales ou des flux en direct.

- Le déploiement du système sur une machine locale.

La stratégie adoptée repose sur une séparation claire entre les phases de traitement de données, de modélisation, d'entraînement et de déploiement, afin de garantir la flexibilité, la reproductibilité et l'évolutivité de notre solution.

2.2 Architecture générale du système:

L'architecture générale du système peut être divisée en trois blocs principaux, interconnectés de manière fluide :

- **Bloc 1** : Prétraitement et traitement des données

Ce bloc comprend l'importation, le redimensionnement et la normalisation des images, ainsi que leur conversion en niveaux de gris. Il prépare les données d'entrée pour le modèle.

- **Bloc 2** : Modèle de classification (CNN)

C'est le cœur du système. Le modèle CNN est entraîné à partir des jeux de données prétraités, puis sauvegardé sous forme de fichier modèle (.h5). Ce modèle est ensuite chargé à chaque prédiction afin de classer l'émotion soit en temps réel via une webcam, soit à partir d'une image statique.

- **Bloc 3** : Interface utilisateur (Frontend)

Cette interface, développée avec Streamlit, permet à l'utilisateur d'activer la webcam ou de charger une image locale. Elle interagit avec le modèle pour fournir la prédiction de l'émotion et l'afficher à l'écran.

Cette architecture permet une séparation des responsabilités tout en assurant une communication efficace entre les composants. De plus, elle facilite l'évolution future du système, par exemple en remplaçant le modèle CNN par un modèle plus avancé ou en adaptant l'interface à un usage web ou mobile.

3. Préparation des données :

Une étape cruciale dans la réussite de tout projet d'apprentissage profond est la préparation rigoureuse des données d'entraînement. La qualité et la diversité des données influencent directement les performances du modèle. Dans le cadre de notre système de reconnaissance des émotions faciales, nous avons utilisé une base de données spécialisée, que nous avons ensuite traitée pour la rendre compatible avec notre modèle convolutionnel.

3.1 Description du dataset utilisé :

Pour entraîner et évaluer le modèle de reconnaissance d'émotions faciales, nous avons utilisé le dataset **FER-2013**, l'un des plus connus et accessibles dans ce domaine. Ce jeu de données a été introduit dans le cadre du concours Kaggle Challenge organisé par ICML en 2013.

Caractéristiques principales du dataset FER-2013 :

- Nombre total d'images : 35 887
- Dimensions : 48x48 pixels
- Type : images en niveaux de gris
- Nombre de classes (émotions) : 7
 - 0 = Colère
 - 1 = Dégoût
 - 2 = Peur
 - 3 = Heureux
 - 4 = Triste
 - 5 = Surprise
 - 6 = Neutre

Les images sont réparties en deux sous-ensembles :

- Entraînement : 28 709 images
- Test : 7178 images

3.2 Prétraitements appliqués aux images :

Avant d'être introduites dans le réseau neuronal, les images brutes subissent plusieurs étapes de prétraitement afin d'assurer une cohérence et d'améliorer la performance du modèle :

- **Redimensionnement** : Toutes les images sont converties au format 48x48 pixels, en cohérence avec la structure initiale du dataset et les exigences du modèle CNN.
- **Conversion en niveaux de gris** : Le modèle CNN s'appuie sur l'intensité des pixels pour extraire les caractéristiques. Ainsi, seules les informations en niveaux de gris sont conservées pour réduire la complexité du traitement.
- **Normalisation** : Les pixels sont normalisés dans l'intervalle [0, 1] en divisant chaque valeur par 255, afin d'accélérer la convergence du modèle.
- **Encodage des labels** : Les étiquettes des émotions sont converties en vecteurs one-hot, facilitant leur traitement lors de l'entraînement avec des fonctions de coût comme la cross-entropy.

Ces étapes de préparation assurent que les données en entrée soient de qualité uniforme, permettant ainsi au réseau de mieux apprendre les motifs caractéristiques des différentes émotions.

4. Conception du modèle CNN :

La reconnaissance des émotions faciales repose sur la capacité d'un réseau de neurones à détecter et à différencier des motifs subtils dans les expressions du visage. Pour ce faire, nous avons conçu une architecture de réseau de neurones convolutifs (CNN) adaptée à notre problématique, avec un bon compromis entre précision et efficacité computationnelle.

4.1. Choix de l'architecture :

Pour détecter efficacement les expressions faciales dans des images en niveaux de gris de petite taille (48x48), nous avons opté pour une architecture de réseau de neurones convolutifs (CNN) profonde et progressive. Notre conception repose sur une succession de blocs convolutifs permettant une extraction hiérarchique des caractéristiques, allant des traits les plus simples (bords, textures) aux motifs complexes caractéristiques des émotions.

Nous avons structuré notre modèle en quatre blocs principaux, chacun composé de deux couches de convolution avec des filtres de **taille 3x3**, suivies d'une normalisation par lot (**Batch Normalization**), d'une couche de regroupement spatial (**MaxPooling**) pour la réduction de dimension, et d'un **dropout** pour limiter le sur-apprentissage.

Après ces blocs, nous avons ajouté une couche entièrement connectée (**Dense**) avec régularisation L2 et dropout, suivie d'une couche de sortie de **7 neurones** avec activation **softmax**, correspondant aux 7 émotions cibles.

Voici la structure détaillée de notre modèle :

- Bloc 1: **Conv2D(64) ×2 → BatchNorm ×2 → MaxPooling → Dropout(0.25)**
- Bloc 2: **Conv2D(128) ×2 → BatchNorm ×2 → MaxPooling → Dropout(0.25)**
- Bloc 3: **Conv2D(256) ×2 → BatchNorm ×2 → MaxPooling → Dropout(0.25)**
- Bloc 4: **Conv2D(512) ×2 → BatchNorm ×2 → MaxPooling → Dropout(0.25)**
- Fully Connected : **Dense(256, ReLU) + L2(0.01) → BatchNorm → Dropout(0.5)**
- Sortie : **Dense(7, softmax)**

Cette architecture nous a permis de capturer les informations discriminantes des expressions faciales tout en assurant la stabilité de l'apprentissage et la capacité de généralisation du modèle.

```

49 # Notre modèle CNN
50 model = Sequential()
51
52 # Bloc 1
53 model.add(Conv2D(64, (3, 3), activation='relu', padding='same', input_shape=(48, 48, 1)))
54 model.add(BatchNormalization())
55 model.add(Conv2D(64, (3, 3), activation='relu', padding='same'))
56 model.add(BatchNormalization())
57 model.add(MaxPooling2D(pool_size=(2, 2)))
58 model.add(Dropout(0.25))
59
60 # Bloc 2
61 model.add(Conv2D(128, (3, 3), activation='relu', padding='same'))
62 model.add(BatchNormalization())
63 model.add(Conv2D(128, (3, 3), activation='relu', padding='same'))
64 model.add(BatchNormalization())
65 model.add(MaxPooling2D(pool_size=(2, 2)))
66 model.add(Dropout(0.25))
67
68 # Bloc 3
69 model.add(Conv2D(256, (3, 3), activation='relu', padding='same'))
70 model.add(BatchNormalization())
71 model.add(Conv2D(256, (3, 3), activation='relu', padding='same'))
72 model.add(BatchNormalization())
73 model.add(MaxPooling2D(pool_size=(2, 2)))
74 model.add(Dropout(0.25))
75
76 # Bloc 4 (facultatif mais booste encore plus)
77 model.add(Conv2D(512, (3, 3), activation='relu', padding='same'))
78 model.add(BatchNormalization())
79 model.add(Conv2D(512, (3, 3), activation='relu', padding='same'))
80 model.add(BatchNormalization())
81 model.add(MaxPooling2D(pool_size=(2, 2)))
82 model.add(Dropout(0.25))
83
84 # Fully Connected
85 model.add(Flatten())
86 model.add(Dense(256, activation='relu', kernel_regularizer=l2(0.01)))
87 model.add(BatchNormalization())
88 model.add(Dropout(0.5))
89 model.add(Dense(7, activation='softmax')) # 7 émotions dans FER-2013
90
91 # === 4. Compilation ===
92 model.compile(
93     optimizer=Adam(learning_rate=0.0001),
94     loss='categorical_crossentropy',
95     metrics=['accuracy']
96 )

```

Figure 3.1 – Construction du modèle CNN profond.

4.2. Hyperparamètres et stratégie d'optimisation :

Pour entraîner efficacement notre modèle, nous avons défini un ensemble d'hyperparamètres adaptés à notre jeu de données et à notre infrastructure matérielle. Les principaux hyperparamètres utilisés sont résumés dans le Tableau 3.1.

Tableau 3.1 - Hyperparamètres choisis pour l'entraînement du modèle CNN.

Hyperpara mètre	Valeur choisie
Optimiseur	Adam

Taux d'apprentissage	0.0001
Fonction de perte	Categorical Crossentropy
Métrique	Accuracy
Batch size	64
Nombre d'époques	30 et 100 (avec arrêt anticipé)
Stratégie d'arrêt	EarlyStopping (patience = 10)
Sauvegarde du modèle	ModelCheckpoint (val_accuracy)

Nous avons utilisé l'optimiseur Adam pour sa capacité à ajuster dynamiquement le taux d'apprentissage pendant l'entraînement. Le faible taux d'apprentissage initial (0.0001) permet une convergence plus stable et progressive.

Afin de prévenir le sur-apprentissage, nous avons intégré deux mécanismes complémentaires :

- EarlyStopping, qui interrompt l'entraînement si la performance en validation cesse de s'améliorer après 10 époques ;
- ModelCheckpoint, qui sauvegarde automatiquement le modèle ayant obtenu la meilleure précision en validation.

Nous avons également utilisé une régularisation L2 sur la couche dense pour pénaliser les poids trop élevés, renforçant ainsi la robustesse du modèle.

5. Développement de l'interface utilisateur :

5.1. Description des fonctionnalités principales :

Dans le but de rendre l'application de reconnaissance des émotions faciales accessible et interactive, nous avons développé une interface utilisateur conviviale permettant deux modes d'utilisation distincts. Ces fonctionnalités sont destinées à démontrer les capacités de notre modèle dans des contextes variés, tout en facilitant l'expérimentation par des utilisateurs non spécialisés.

5.1.1. Mode reconnaissance en temps réel (webcam) :

Dans ce mode, l'application capture en continu les images du visage à partir de la webcam de l'utilisateur. Chaque image capturée est automatiquement traitée par le système qui :

- Détecte les visages à l'aide de l'algorithme Haar Cascade de OpenCV.
- Redimensionne et normalise les régions faciales détectées.
- Prédit l'émotion dominante parmi les sept catégories apprises par le modèle (colère, dégoût, peur, bonheur, tristesse, surprise, neutre).
- Affiche le résultat en temps réel sur la vidéo, avec une boîte englobante et une étiquette indiquant l'émotion reconnue.

Ce mode permet une interaction immédiate avec le système et illustre la capacité du modèle à s'adapter aux variations d'expression dans un contexte dynamique.

5.1.2. Mode classification à partir d'images

Ce second mode est destiné à tester le système sur des images fixes chargées manuellement par l'utilisateur via l'interface. Le processus comprend :

- L'Importation d'une image depuis l'ordinateur de l'utilisateur.
- La détection automatique des visages dans l'image.
- Le traitement de chaque visage détecté, incluant redimensionnement, conversion en niveaux de gris et normalisation.
- La prédiction de l'émotion faciale et l'affichage du résultat.

Ce mode est particulièrement utile pour tester le système sur des bases d'images spécifiques ou des cas d'usage précis.

5.2 Intégration des éléments d'évaluation dans l'interface :

Afin de permettre une compréhension approfondie du comportement du modèle, nous avons intégré dans l'interface utilisateur une fonctionnalité optionnelle permettant d'afficher les résultats de l'évaluation. Cette fonctionnalité est accessible à l'aide d'une case à cocher (checkbox) dans la barre latérale de l'application.

Lorsque l'utilisateur active cette option, plusieurs éléments sont affichés dans la zone principale de l'interface :

a. Courbe d'apprentissage

Nous présentons une courbe illustrant l'évolution de la précision (accuracy) sur les ensembles d'entraînement et de validation au fil des époques. Cette visualisation permet d'observer la convergence du modèle, ainsi que les éventuels signes de sur-apprentissage (overfitting).

b. Matrice de confusion

Une matrice de confusion générée à partir de l'ensemble de test est affichée sous forme d'image. Elle permet d'identifier quelles émotions sont correctement reconnues et où se situent les confusions les plus fréquentes.

c. Métriques textuelles

Nous listons également les principaux paramètres d'apprentissage utilisés (fonction de perte, fonction d'activation, optimiseur) ainsi que les indicateurs de performance finaux, tels que :

- **La précision globale (accuracy).**
- **Le F1-score.**
- **La précision (precision),**
- **Le rappel (recall),**

Ces métriques sont extraites à partir du rapport de classification sauvegardé dans un fichier texte au format lisible, et affichées directement dans l'interface à l'aide d'un composant de type `st.code()`.

Cette section optionnelle d'évaluation offre aux utilisateurs ou aux examinateurs une meilleure visibilité sur la qualité du modèle, tout en gardant l'interface principale simple et focalisée sur la prédiction. Elle s'adresse principalement aux utilisateurs techniques souhaitant explorer les performances internes du modèle.

5.3 Technologies utilisées (Python, TensorFlow/Keras, OpenCV, etc.) :

La mise en œuvre de notre système de reconnaissance des émotions faciales repose sur l'utilisation de plusieurs technologies complémentaires, soigneusement sélectionnées pour leur robustesse, leur compatibilité et

leur richesse fonctionnelle. Ci-dessous, nous présentons les principaux outils et bibliothèques utilisés tout au long du projet.

a. Python

Le langage Python a été utilisé comme socle principal pour le développement du projet. Sa syntaxe simple, sa large communauté et son écosystème riche en bibliothèques dédiées à l'intelligence artificielle en font un choix privilégié pour les projets de deep learning.

b. TensorFlow et Keras

La conception, l'entraînement et l'évaluation du modèle de réseau de neurones convolutifs (CNN) ont été réalisés à l'aide de la bibliothèque TensorFlow, via son API haut niveau Keras. Ces outils nous ont permis de :

- Définir l'architecture du modèle de manière modulaire,
- Utiliser des techniques avancées comme l'arrêt anticipé (EarlyStopping) et la sauvegarde du meilleur modèle (ModelCheckpoint).
- Monitorer facilement les métriques pendant l'entraînement.

c. OpenCV

OpenCV a été utilisé pour le traitement d'images, notamment pour :

- La capture vidéo depuis la webcam.
- La détection des visages avec les classifieurs en cascade de Haar.
- La conversion en niveaux de gris et le redimensionnement des images en entrée du modèle.

d. Streamlit

Pour l'interface utilisateur, nous avons choisi Streamlit, un framework open source permettant de créer rapidement des applications web interactives à partir de scripts Python. Grâce à Streamlit, nous avons pu :

- Créer une interface intuitive pour lancer la détection en direct ou charger des images.

- Intégrer des visualisations (images, graphiques, matrices de confusion).
- Afficher dynamiquement les résultats des prédictions.

e. Autres bibliothèques

- **NumPy** : pour la manipulation efficace des tableaux de données.
- **Pillow (PIL)** : pour le chargement et l'affichage d'images dans l'interface.
- **Matplotlib & Seaborn** : pour la génération des courbes d'apprentissage et de la matrice de confusion.
- **scikit-learn** : pour le calcul des métriques de classification (précision, recall, F1-score).

L'intégration de ces outils a permis de construire une application performante, interactive et facilement exploitable, tant pour l'entraînement du modèle que pour son déploiement.

```
train_model.py > ...
1 import os
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from tensorflow.keras.preprocessing.image import ImageDataGenerator
5 from tensorflow.keras.models import Sequential
6 from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
7 from tensorflow.keras.optimizers import Adam
8 from tensorflow.keras.callbacks import ModelCheckpoint
9 from tensorflow.keras.callbacks import EarlyStopping
10 from tensorflow.keras.layers import BatchNormalization
11 from tensorflow.keras.regularizers import l2
```

Figure 3.2 – Importation des bibliothèques nécessaires à l'entraînement du modèle.

```
evaluate_model.py > ...
1 from tensorflow.keras.models import load_model
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from tensorflow.keras.preprocessing.image import ImageDataGenerator
5 from sklearn.metrics import classification_report, confusion_matrix
6 import seaborn as sns
```

Figure 3.3 – Importation des bibliothèques nécessaires à l'évaluation du modèle.

```

app.py > ...
1  import streamlit as st
2  import numpy as np
3  import cv2
4  from tensorflow.keras.models import load_model
5  from PIL import Image
6  import time

```

Figure 3.4 – Importation des bibliothèques nécessaires à l'application FER.

6. Entraînement et validation du modèle :

L'entraînement du modèle CNN a été réalisé à partir de notre jeu de données prétraité, en utilisant des techniques d'optimisation et de régularisation visant à maximiser la performance tout en limitant le sur-apprentissage. Cette section présente les métriques d'évaluation utilisées ainsi que les résultats obtenus après la phase d'entraînement.

```

train_model.py > ...
99  # Entraînement
100  checkpoint = ModelCheckpoint("model/fer_cnn_model.h5", save_best_only=True, monitor='val_accuracy', mode='max')
101
102  early_stop = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
103
104
105  history = model.fit(
106      train_generator,
107      epochs=100,
108      validation_data=test_generator,
109      callbacks=[checkpoint, early_stop]
110  )

```

Figure 3.5 – Entraînement du modèle.

```

evaluate_model.py > ...
7
8 #Chemin vers les données
9 test_dir = "data/test"
10
11 # Préparation des données de test
12 test_datagen = ImageDataGenerator(rescale=1./255)
13 test_generator = test_datagen.flow_from_directory(
14     test_dir,
15     target_size=(48, 48),
16     color_mode='grayscale',
17     batch_size=64,
18     class_mode='categorical',
19     shuffle=False
20 )
21
22 # Charger le modèle
23 model = load_model('model/fer_cnn_model.h5')
24
25 # Générer les prédictions
26 predictions = model.predict(test_generator, verbose=1)
27 predicted_classes = np.argmax(predictions, axis=1)
28
29 # Vrais labels et noms des classes
30 true_classes = test_generator.classes
31 class_labels = list(test_generator.class_indices.keys())

```

Figure 3.6 – Chargement du modèle et génération des prédictions sur l'ensemble de test.

6.1. Métriques d'évaluation utilisées :

Afin d'évaluer les performances du modèle de manière fiable et complète, nous avons utilisé les métriques suivantes :

- **Accuracy (précision globale)** : proportion d'exemples correctement classés parmi l'ensemble du jeu de test.
- **Matrice de confusion** : permet de visualiser les prédictions correctes et incorrectes par classe, offrant une vue détaillée des performances par émotion.
- **Précision (Precision)** : capacité du modèle à éviter les faux positifs pour chaque classe.
- **Rappel (Recall)** : capacité du modèle à détecter tous les exemples pertinents (réduction des faux négatifs).

- **F1-score** : moyenne harmonique entre la précision et le rappel, utile en cas de classes déséquilibrées.

Ces métriques ont été calculées à l'aide des fonctions de la bibliothèque **scikit-learn**, et les résultats ont été enregistrés sous forme de rapport textuel ainsi que sous forme graphique.

6.2. Résultats obtenus :

L'entraînement du modèle a été réalisé sur 30 puis 100 époques, avec un arrêt anticipé (patience = 10), en utilisant un lot de 64 images (batch size = 64). Les deux graphiques ci-dessous illustrent l'évolution de la précision sur les ensembles d'entraînement et de validation.

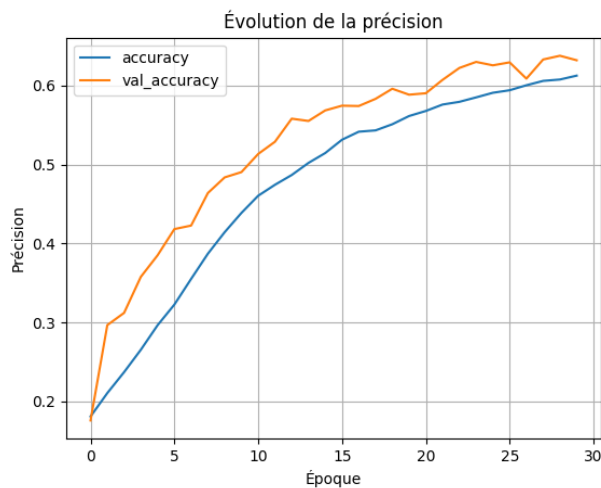


Figure 3.7 – Courbe d'apprentissage du modèle pour 30 époques.

Comme illustré, le modèle converge rapidement dès les premières époques, avec une amélioration stable de la précision jusqu'à atteindre un plateau autour de la 20e époque.

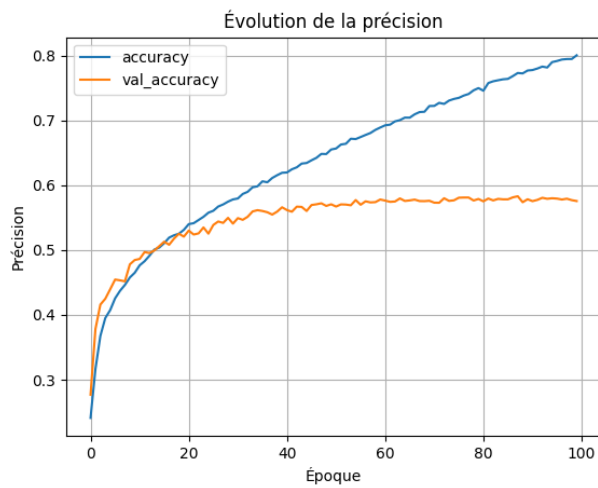


Figure 3.8 – Courbe d’apprentissage du modèle pour 100 époques.

Le modèle final a atteint une précision globale(**Accuracy**) : 80 %

Une matrice de confusion a également été générée pour évaluer les performances par émotion :

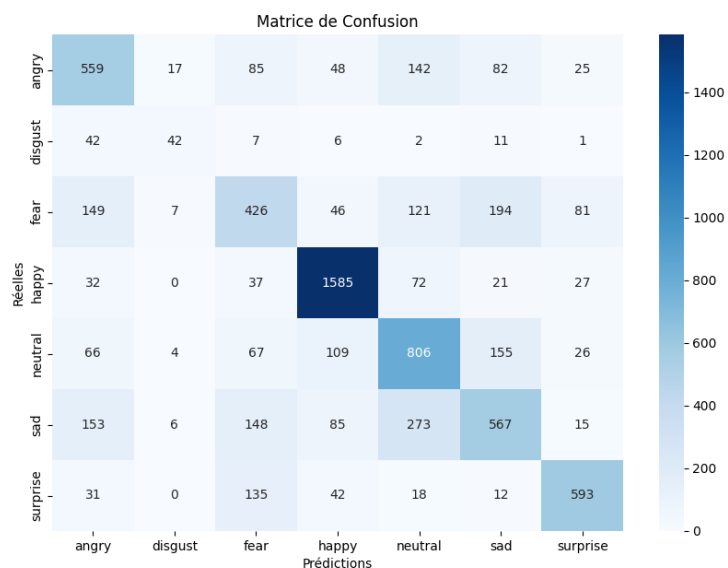


Figure 3.9 – Matrice de confusion du modèle

L'analyse de cette matrice révèle que certaines émotions comme "Happy" ou "Neutral" sont bien reconnues, tandis que d'autres comme "Fear" et "Disgust" sont plus fréquemment confondues, ce qui est cohérent avec les similarités visuelles entre ces expressions.

Extrait du rapport de classification :

precision	recall	f1-score	support	
angry	0.54	0.58	0.56	958
disgust	0.55	0.38	0.45	111
fear	0.47	0.42	0.44	1024
happy	0.83	0.89	0.86	1774
neutral	0.56	0.65	0.60	1233
sad	0.54	0.45	0.50	1247
surprise	0.77	0.71	0.74	831
accuracy			0.64	7178
macro avg	0.61	0.58	0.59	7178
weighted avg	0.63	0.64	0.63	7178

Figure 3.10 - rapport de classification.

Ces résultats montrent que le modèle est capable de généraliser efficacement sur des données jamais vues, tout en maintenant un bon équilibre entre performance et complexité.

7. Déploiement du système :

Afin de rendre notre modèle de reconnaissance des émotions faciales accessible et utilisable dans un cadre applicatif, nous avons procédé à son déploiement sous la forme d'une interface utilisateur interactive. Cette application, développée avec Streamlit, permet à l'utilisateur de tester le modèle en temps réel à l'aide d'une webcam ou à partir d'images statiques, tout en accédant aux métriques d'évaluation du modèle.

Figure 3.11 présente l'interface d'accueil de l'application, qui constitue le point d'entrée pour l'utilisateur. Une version alternative en mode sombre est disponible pour le confort visuel de l'utilisateur, comme illustré dans la Figure 7.2.

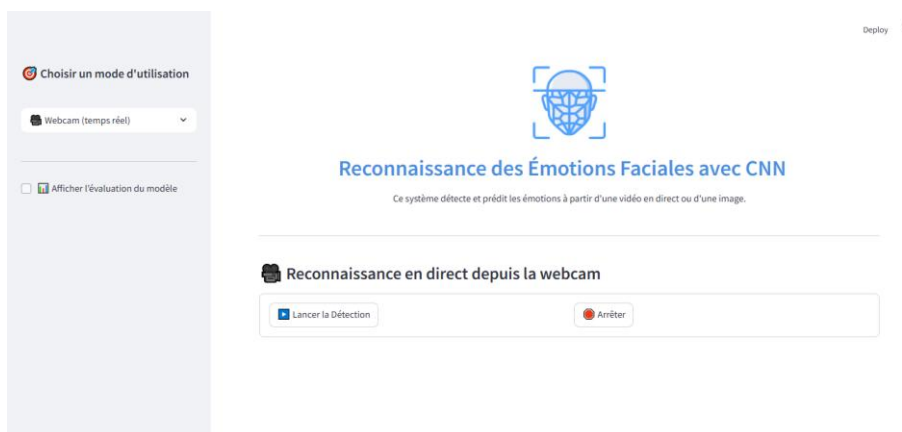


Figure 3.11- Interface d'accueil de l'application.



Figure 3.12- Interface d'accueil – Mode sombre

7.1. Détection des visages en temps réel

La détection en temps réel constitue une fonctionnalité essentielle du système. Grâce à la bibliothèque OpenCV, l'application est capable d'accéder à la webcam de l'utilisateur, de capturer le flux vidéo et de détecter les visages à l'aide du classifieur Haar Cascade. Une fois le visage localisé, la région d'intérêt est extraite, prétraitée (mise en niveaux de gris, redimensionnement, normalisation) puis envoyée au modèle pour prédiction.

L'affichage est dynamique : l'émotion détectée est superposée directement sur la vidéo en temps réel, permettant à l'utilisateur de visualiser instantanément les résultats de la reconnaissance.

Cette fonctionnalité a été testée sur un PC doté d'un processeur Intel i3-6006U avec 8 Go de RAM, et s'est révélée fluide, avec un temps de réponse satisfaisant malgré des ressources limitées.

7.2. Prédiction de l'émotion (temps réel + image statique)

L'application propose deux modes d'utilisation :

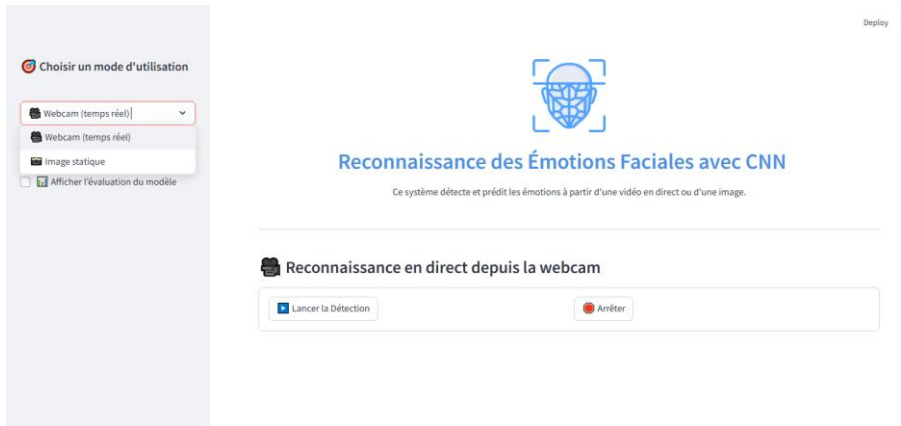


Figure 3.13- Interface de sélection des deux modes de détection.

- **Mode image statique** (Figure 3.14) : l'utilisateur peut charger une image locale contenant un visage (Figure 3.15). L'image est prétraitée de manière identique à celle du flux vidéo, puis soumise au modèle CNN. L'émotion détectée est ensuite affichée à côté de l'image importée (Figure 3.16).

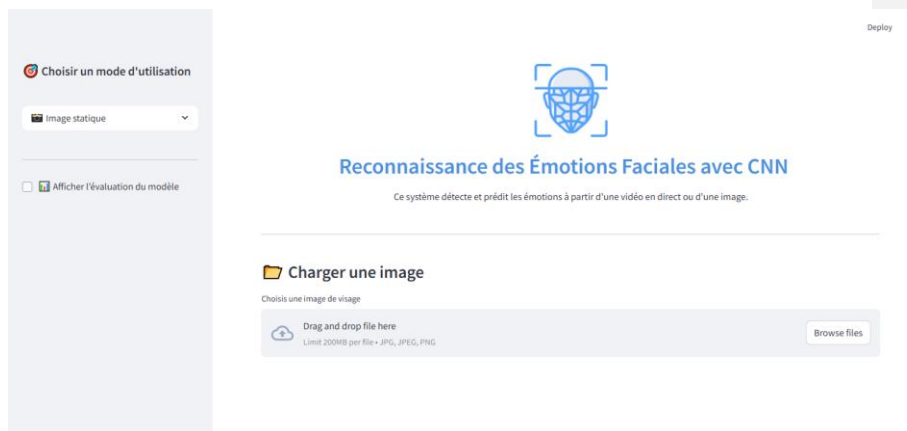


Figure 3.14- Interface du mode - Image statique.

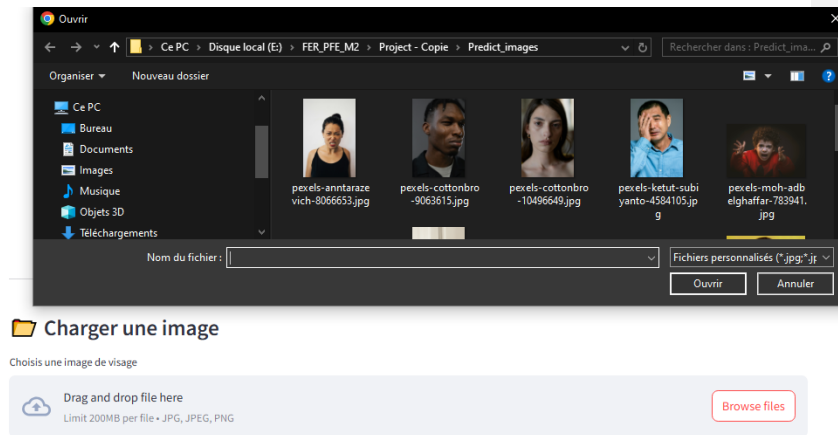


Figure 3.15- Chargement d'une image par l'utilisateur.

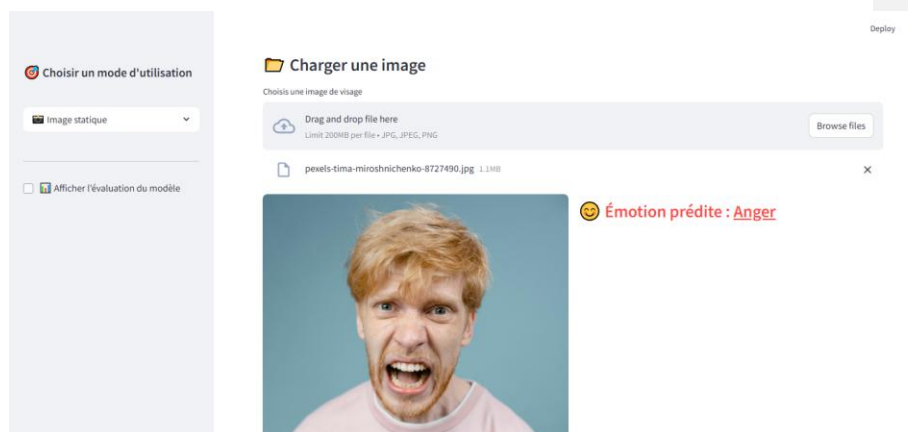


Figure 3.16- Exemple de prédiction d'émotion sur une image statique.

- **Mode webcam (temps réel)** : (Figure 3.17) comme décrit précédemment, le flux vidéo est traité en continu, chaque visage détecté étant analysé image par image pour en prédire l'émotion.



Figure 3.17- Interface du mode - Webcam temps réel.

L'utilisateur peut lancer la détection via un bouton dédié "Lancer la détection" (Figure 3.18), et l'arrêter à tout moment à l'aide du bouton d'arrêt (Figure 3.19). Le système traite en continu les flux vidéo, détecte les visages présents dans chaque image et affiche l'émotion prédite en surimpression.

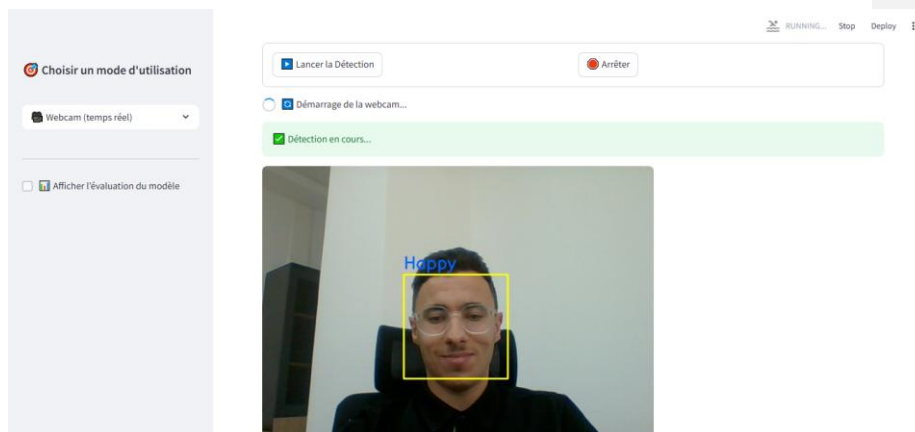


Figure 3.18- Lancement de la détection des émotions via webcam.

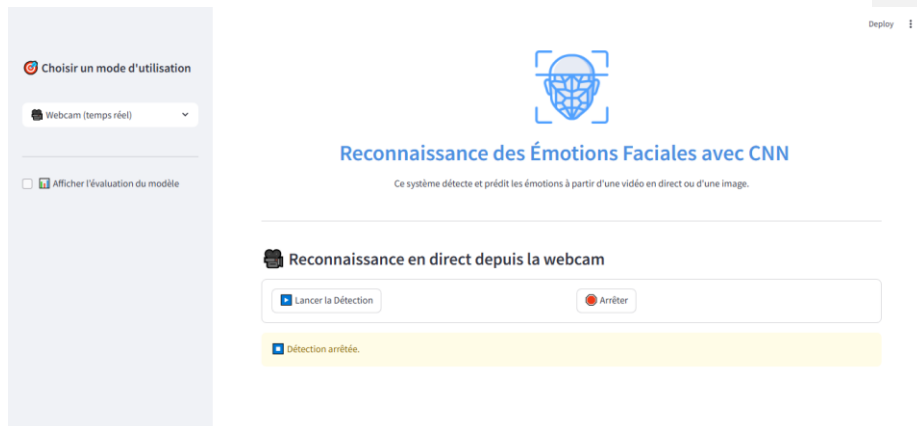


Figure 3.19- Arrêt de la détection des émotions via webcam.

Une troisième option, accessible via une case à cocher dans la barre latérale, permet d'afficher les résultats d'évaluation du modèle :

- **Courbe d'apprentissage** (Figure 3.20) (précision en entraînement et validation)

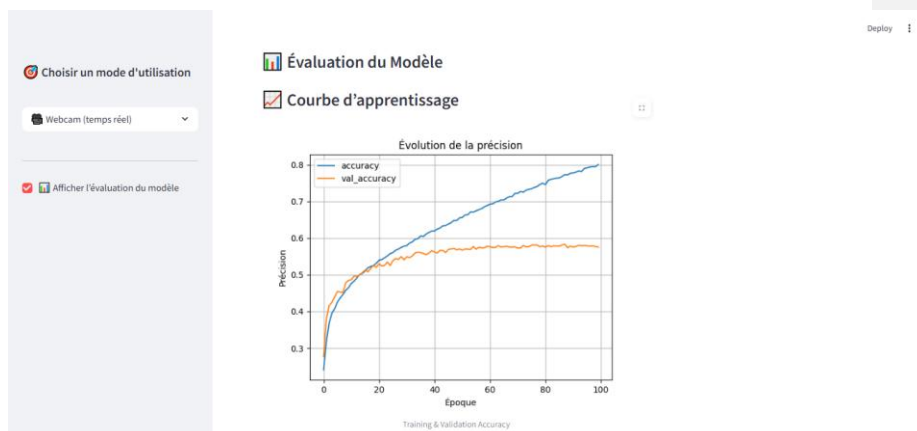


Figure 3.20- Courbe d'apprentissage du modèle CNN.

- **Matrice de confusion** (Figure 3.21)

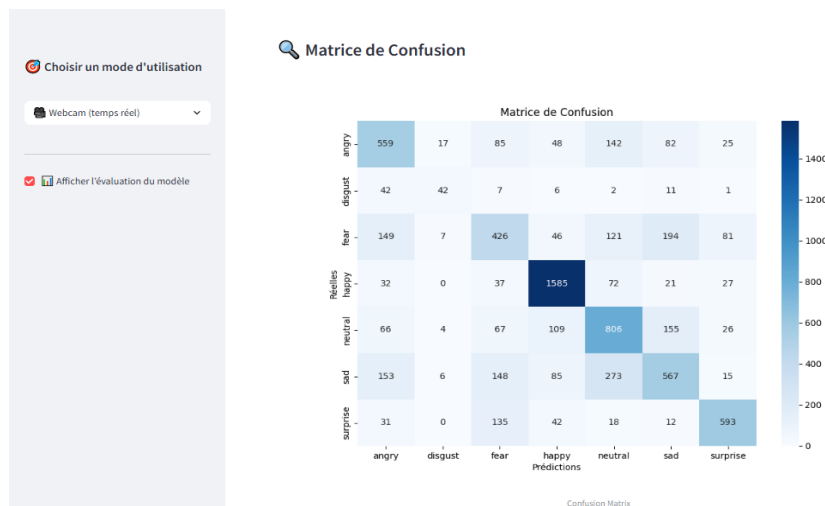


Figure 3.21- Matrice de confusion des émotions détectées.

- **Rapport de classification** avec les scores de précision, rappel et F1-score (Figure 3.22).



Figure 3.22- Rapport de classification : précision, rappel et F1-score.

Cette approche rend l'application complète, à la fois pour la démonstration du fonctionnement du modèle et pour l'analyse de ses performances globales.

8. Analyse critique des résultats :

L'évaluation du modèle de reconnaissance des émotions faciales révèle des performances globalement satisfaisantes dans le cadre du jeu de

données utilisé et des objectifs du projet. Toutefois, une analyse critique est essentielle pour mieux comprendre les points forts et les limites du système, ainsi que les pistes potentielles d'amélioration.

8.1. Discussion sur les performances :

Le modèle CNN a atteint une précision de validation de 80 %, ce qui témoigne d'une bonne capacité à généraliser les apprentissages sur des données nouvelles. La courbe d'apprentissage montre une progression stable sans sur-apprentissage, notamment grâce à l'utilisation du mécanisme d'arrêt anticipé (EarlyStopping) et à la régularisation L2.

La matrice de confusion met en évidence des performances variables selon les émotions. Par exemple, les émotions comme la joie (Happy) et la surprise (Surprise) sont généralement bien reconnues, tandis que des émotions telles que la peur (Fear), le dégoût (Disgust) ou la colère (Anger) présentent davantage de confusion, notamment entre elles. Cela peut s'expliquer par :

- La similarité visuelle entre certaines expressions faciales ;
- Un déséquilibre dans le nombre d'exemples par classe dans le jeu de données ;
- Les limites des représentations 2D d'expressions faciales (notamment en cas d'ombres ou de variations d'angle).

Malgré ces défis, les scores de F1-score et de rappel sont jugés satisfaisants pour un modèle déployé dans un environnement standard avec des ressources limitées.

9. Conclusion :

Dans ce chapitre, nous avons présenté en détail l'ensemble du processus de développement de notre système de reconnaissance des émotions faciales basé sur les réseaux de neurones convolutifs (CNN). Ce processus a été structuré selon une approche rigoureuse et progressive, allant de la préparation des données jusqu'à l'évaluation finale du modèle.

Nous avons commencé par décrire la structuration du jeu de données et les différentes étapes de prétraitement nécessaires à l'entraînement d'un modèle efficace. L'architecture CNN que nous avons conçue a été choisie pour son équilibre entre complexité, précision et faisabilité sur une machine à ressources limitées. Grâce à des techniques comme la régularisation, le dropout, l'arrêt anticipé et l'optimisation avec Adam, nous avons pu obtenir un modèle stable et performant.

L'interface utilisateur, développée avec Streamlit, permet une interaction intuitive en proposant deux modes d'utilisation : la reconnaissance d'émotions à partir d'images statiques et celle en temps réel via webcam. Une fonctionnalité supplémentaire d'affichage des métriques d'évaluation a été intégrée, renforçant l'aspect pédagogique et démonstratif du projet.

Les résultats obtenus montrent une précision satisfaisante dans l'identification des émotions, avec quelques confusions entre certaines classes proches. Cette observation a conduit à une analyse critique dans la dernière section, où nous avons identifié plusieurs limites du système ainsi que des pistes d'amélioration prometteuses.

Ce chapitre constitue ainsi la colonne vertébrale technique du projet, synthétisant la démarche méthodologique mise en œuvre pour concevoir un système fonctionnel, exploitable et potentiellement évolutif dans divers contextes d'application.

Conclusion générale

Ce mémoire a porté sur la conception et la mise en œuvre d'un système de reconnaissance des émotions faciales basé sur les réseaux de neurones convolutifs (CNN). L'objectif principal était de concevoir un modèle capable d'identifier automatiquement les expressions faciales à partir d'images statiques ou de flux vidéo en temps réel, en s'appuyant sur des techniques modernes d'intelligence artificielle et de vision par ordinateur.

Notre modèle CNN entraîné sur un ensemble d'images d'émotions faciales a démontré une précision globale de **80** qui satisfaisante dans la classification, avec des résultats prometteurs sur des émotions comme la joie, la surprise ou la neutralité.

L'interface utilisateur développée avec Streamlit offre deux modes de fonctionnement intuitifs : la prédiction à partir d'images statiques et la reconnaissance en temps réel via webcam.

Notons que les courbes d'apprentissage, la matrice de confusion et les métriques d'évaluation (précision, rappel, F1-score) ont confirmé la stabilité du modèle ainsi que sa capacité à généraliser sur des données de test.

Le système proposé illustre le potentiel des réseaux de neurones profonds dans des applications concrètes de détection et d'analyse du comportement humain. Ce travail contribue à montrer qu'il est possible de concevoir un système de reconnaissance d'émotions fiable avec des moyens limités, ce qui ouvre la voie à des applications dans la sécurité, l'éducation, les interfaces homme-machine ou encore la santé mentale.

Sur le plan académique, ce projet constitue un exemple d'intégration réussie entre des concepts théoriques issus du deep learning et leur implémentation dans un environnement logiciel accessible, ce qui ouvre des perspectives intéressantes pour la recherche comme pour l'innovation technologique.

Bibliographie :

- [1] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, 1978.
- [2] R. Plutchik, *The Emotions: Facts, Theories, and a New Model*. New York, NY: Random House, 1980.
- [3] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [4] A. Jack, C. B. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7241–7244, 2012.
- [5] G. Amato, F. Falchi, C. Gennaro, and C. Vairo, "A Comparison of Face Verification with Facial Landmarks and Deep Features," in Proc. Tenth Int. Conf. Advances in Multimedia (MMEDIA), Athens, Greece, Apr. 2018.
- [6] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in Eur. Conf. Computer Vision, 2004.
- [7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2005.
- [8] T. Barbu, "Gabor Filter-Based Face Recognition Technique," Proc. Romanian Academy, Series A, vol. 11, no. 3, pp. 277–283, 2010.
- [9] R. He, X. Wu, Z. Sun, and T. Tan, "Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1761–1773, 2019.
- [10] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014.
- [11] Blog du Modérateur, "27 générateurs d'images par intelligence artificielle," 2025. [Online]. Available: <https://www.blogdumoderateur.com/tools/design/generateur-images-ia/>
- [12] J. Miah et al., "Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks," arXiv:2310.18237, 2023.

[13] *Le Monde*, “Dépistage du cancer du sein : l’intelligence artificielle doit trouver sa juste place,” Apr. 4, 2024. [Online]. Available: https://www.lemonde.fr/sciences/article/2024/04/04/depistage-du-cancer-du-sein-l-intelligence-artificielle-doit-trouver-sa-juste-place_6225918_1650684.html

[14] *Le Monde*, “L’IA ouvre de nouvelles possibilités de surveillance,” May. 23, 2024. [Online]. Available: https://www.lemonde.fr/idees/article/2024/05/23/l-ia-ouvre-de-nouvelles-possibilites-de-surveillance_6235057_3232.html

[15] M. Pantic and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: The state of the art,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, 2000.

[16] M. Bartlett et al., “Real- time face detection and facial expression recognition: Development and application to human- computer interaction,” *Proc. IEEE*, 2000.

[17] Paul Ekman Group, “Universal Facial Expressions Photo Set,” 2006. [Online]. Available: <https://www.paulekman.com/product/universal-facial-expression-pictures-photo-set/>

[18] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, 1986.

[21] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556*, 2014.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proc. NIPS*, pp. 1097–1105, 2012.

[23] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [25] B. Yu and D. Zhang, "Image-based static facial expression recognition with multiple deep network learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015.
- [26] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *CVPR Workshop*, 2010.
- [27] A. Mollahosseini et al., "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affective Comput.*, vol. 10, no. 1, pp. 18–31, 2019.
- [28] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *WACV*, 2016.
- [29] B. Hasani and M. H. Mahoor, "Facial expression recognition using enhanced deep 3D convolutional neural networks," in *CVPRW*, 2017.
- [30] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order," *Pattern Recognit.*, vol. 61, pp. 610–628, 2017.
- [31] B. Charroux and A. Osmani, *UML2 Pratique de la Modélisation*, 2ème éd., Pearson Education France, 2009.
- [32] ResearchGate, "Performance of computer vision models on the ImageNet challenge over time (2010–2017)," en ligne. [Disponible sur : https://www.researchgate.net/figure/Performance-of-computer-vision-models-on-the-ImageNet-challenge-over-time-2010-2017_fig1_354827989]

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة عبد الحميد بن باديس-مستغانم-
كلية العلوم الدقيقة والإعلام الآلي
تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية
لإنجاز البحث

أنا الممضي أدناه،

الطالب(ة): بوطالبي محمد الياس رقم التسجيل

الجامعي: UN27012024202037033751

الحامل لبطاقة التعريف الوطنية رقم: 406826519 والصادرة بتاريخ: 05.09.2023

عن حجاج-مستغانم

المسجل ب: كلية العلوم الدقيقة والإعلام الآلي / قسم : رياضيات وإعلام آلي

شعبة : إعلام آلي / التخصص : الذكاء اصطناعي و انترنت الأشياء

والمكلف بإنجاز مذكرة ماستر بعنوان: التعرف على الانفعالات الوجهية باستخدام

الشبكات العصبية الالتفافية

أُصرح بشرفي أنني ألتزم بمراعاة المعايير العلمية والمنهجية ومعايير الأخلاقيات العلمية

والنزاهة الأكاديمية المطلوبة في إنجاز البحث ، وأتحمل المسؤولية الشخصية عن كل

المحتوى المتضمن في البحث المذكور أعلاه .

التاريخ: 28/06/2025

إمضاء المعني



* ملحق القرار الوزاري رقم 933 المؤرخ في 28 جويلية 2016 الذي يحدد القواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.